STUDIES IN UTILITY THEORY

by

STIG OWE LARSSON
B.Sc.(Hons. with distinction)
Sir George Williams University, 1970
M.Sc., University of Alberta, 1973

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES
THE FACULTY OF COMMERCE AND BUSINESS ADMINISTRATION

We accept this thesis as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA
October, 1977

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the Head of my Department or by his representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Commerce

The University of British Columbia
2075 Wesbrook Place
Vancouver, Canada
V6T 1W5

Date April 7 1978

ABSTRACT

Since vonNeumann and Morgenstern made their contributions, the expected utility criterion (EUC) has been the most accepted criterion in decision theory. Following their axiomatic approach justifying EUC, several other studies have been made suggesting the same criterion but under slightly different axiomatic systems. However, critics have found several simple decision problems (called paradoxes) which seem to contradict the conclusions of EUC; that is, the paradoxes contradict one or more of the axioms made to support EUC. The criticisms are based on empirical studies made in regard to the paradoxes. It is not always obvious, however, which axiom(s) is not accepted, since each approach to EUC gives a set of sufficient rather than necessary assumptions for EUC to hold.

In Part I of the thesis a set of axioms which are necessary for EUC to hold is specified. Each of these axioms contains a basic assumption of a decision maker's behaviour. Therefore by considering the paradoxes in terms of these axioms, a better understanding is obtained with regard to which properties of EUC seem to be contradicted by the paradoxes.

The conclusion of this study shows that most people contradict EUC because it does not differentiate between a "known" risk and an "unknown" risk. In Knight's terminology, there is a distinction between decision making under risk and uncertainty. Most empirical studies show that these differences are of such substantial proportions that there is a questionable justification for using the expected utility criterion for decision making under uncertainty. Although many alternatives

to EUC for decision making under uncertainty exist, there are very few criteria for decision problems which fall between risk and uncertainty, that is, partial risk problems. Those existing are of an ad hoc nature. As a normative theory the EUC is far superior to any of these criteria in spite of its lack of distinction between risk and uncertainty.

In the second part of the thesis an alternative normative criterion is suggested for decision making under partial risk and uncertainty. As an extension of EUC, this criterion distinguishes between risk and uncertainty. This theory expands on Ellsberg's suggestion that "ambiguity" influences one's preference among a set of alternatives. In this extension a more precise definition of "ambiguity" is needed and one is suggested here as a relation on the inner and outer measure of an event. The extension of EUC is then obtained by considering a more general set function, termed P-measure, which would depend on a set's ambiguity rather than a probability measure on the sets of rewards. It is concluded by an axiomatic development that the P-measure must be a non-negative monotonic set function which is not necessarily additive. It is also shown that the standard paradoxes related to paradoxes based on "known" versus "unknown" probabilities may be explained by this method and would therefore suggest an alternative to EUC for decision making under partial risk and uncertainty.

Table of Contents

Acknowledgements

I gratefully acknowledge my great debt to my committee for their work in making this thesis possible.

I am grateful to Dr. K. MacCrimmon, my thesis supervisor jointly with Dr. S. Brumelle, for his guidance in the decision theory in both the selection of the problems of interest and in the recommendation of relevant papers in the area. I am also very grateful to Dr. S. Brumelle for his interest and suggestions during the development of the thesis and for his careful checking into the consistency of the theory. I am also grateful to Drs. D. Wehrung and W. Ziemba for their careful reading of the material and for their suggested changes to improve it. I am grateful for the discussions with Dr. P. Greenwood in the early stages of the thesis which led up to its present form.

Aside from my committee, I would like to express my gratitude to Dr. M. Chew who faithfully read the thesis, and suggested necessary changes, both in style and content.

INTRODUCTION

The problem of how to make the best choice among some alternatives has interested philosophers, statisticians, politicians, and mathematicians for a long time mainly because everyone is faced constantly with this problem. Even by refusing to choose a choice is made. The theory or science of studying "how to choose" or "what to choose" is called the "Theory of Decision Making".

There exist, of course, subdivisions of the Theory of Decision Making, and the first assumptions we shall make are the following:

i) the set of possible choices is known. We shall denote this set by A, and call it an action set.

ii) the set of consequences of our choices is known. The exact consequences which will occur may not be known for any given alternative. This set will be denoted by R, and called a reward set.

In the first part of the thesis, we shall also make the following assumption:

iii) there exist a set of probabilities (Π) such that for a given action a (belonging to A), we may receive the rewards $r_1,\ldots,r_n$ with probabilities $p_1,\ldots,p_n$ where $p_1+\ldots+p_n = 1$.

It is assumed that an ordering exists on A; that is, we can specify a preference among the alternatives which belong to A.

The theory of decision making under uncertainty is concerned in part with what properties the ordering on A ought to have. For example if we are asked to state our preference ordering for alternatives a, b and c, is it reasonable to have the preference a to b and b to c but c to a? If our preferences were in the order a, b, c, then if alternative c were not offered would it be reasonable to assume that b is preferred to a?

In nearly all Theories of Decision Making it is assumed that the ordering on A can also be specified by a real-valued function f on A. We shall call this function an evaluation function. That is, if alternative a is preferred to alternative b then f(a) is larger than f(b). Therefore the questions "How to choose" and "What to choose" are equivalent to specifying the properties f is assumed to have.

One such function is specified by the expected utility criterion. This criterion specifies two conditions on the ordering on A. Firstly, if a and b are two alternatives belonging to A such that a results in outcome r_1 for certain and b results in outcome r_2 for certain, then there exists a real valued function U on R whose numerical values have the same ordering as the preference ordering on A. The function U will be called a utility function. Secondly, if it is assumed that alternative c gives the rewards $r_1, \ldots, r_n$ with probabilities $p_1, \ldots, p_n$ respectively, then the function f on A is defined by the expected value of the utility function, that is

$$f(c) = p_1U(r_1) + p_2U(r_2) + \ldots + p_nU(r_n).$$

There exist at least two concepts of the meaning of a "utility function" in the economic literature. The first one starts by assuming an ordering on the reward set where the reward set is usually defined as the real line, and the ordering is defined by the natural ordering on the real line. Any increasing function on the real line is, therefore, order preserving. If action a gives rise to the probability measure $P(\cdot,a)$ on the real line, and similarly action b the probability measure $P(\cdot,b)$, an assumption is then made that there exists an order-preserving function U such that a is preferred to b if

$$\int U(r)\,dP(r,a) \geqslant \int U(r)\,dP(r,b).$$

As early as 1738 Bernoulli suggested that $U(r) = \log r$. Another example that has some appeal is the following:

$$U(r) = \begin{cases} 1 & r \geqslant \alpha \\ 0 & r < \alpha \end{cases}$$

where α is any real number. Then, a is preferred to b if

$$P(r \geqslant \alpha, a) \geqslant P(r \geqslant \alpha, b).$$

That is, action a is preferred to action b if the probability of receiving at least α is greater for action a than for action b.

Other examples assume that U is a differentiable function such that

$$\frac{dU}{dr} > 0, \frac{d^2U}{dr^2} < 0.$$

The first condition assumes that the utility increases with wealth, and the second condition assumes that for a fixed increase in wealth our utility decreases as our wealth is increased. For example, log r satisfies these conditions. For an analysis of some of the more specialized utility functions, see Pratt (1964).

This method assumed, therefore, an ordering on R and extended it to A by expected utility, by directly assuming the existence of the utility function.

The second approach to expected utility theory was suggested independently by Ramsey (1926) and vonNeumann-Morgenstern (1947) by showing that the expected utility criterion can be justified on the basis of a set of relatively simple assumptions or axioms on the decision maker's ordering on A. That is, if each one of the axioms is accepted as reasonable, then they jointly imply that the ordering must satisfy the expected utility criterion. The reason for considering a set of axioms implying the expected utility criterion is that hopefully the axioms are simple enough for the decision maker to realize all their implications. He can, therefore, determine whether his preference ordering can be specified by the expected utility criterion.

In this case, however, the utility function is derived for each decision maker, and additional assumptions on the utility function can usually not be made. For example, if a utility function is derived based on Savage's (1954) set of axioms, it is impossible for U to also satisfy

$$\frac{d^2U}{dr^2} < 0$$

since U must be bounded. The difficulties with this approach are that in most cases it is impossible to realize all the implications of any one axiom. Many other authors have since developed their own sets of axioms implying the expected utility criterion in the hope that their set of axioms might be more convincing.

In contrast, the critics of the expected utility criterion have suggested relatively simple decision problems for which many individuals' preferences contradict the criterion. Because of this apparent contradiction these problems are usually called "paradoxes". From some of the "paradoxes" it is not obvious whether expected utility as a criterion is being rejected, or simply one of the suggested axioms. That is, the axioms suggested are sufficient rather than necessary for the expected utility criterion to hold. Clearly the set of possible axioms one can specify as sufficient for the expected utility criterion to hold is very large and we must therefore assess the merit of one set over another.

In Part I of the thesis the main concern will be: what does the expected utility criterion imply of our preference among the alternatives, or, in mathematical terms, we shall consider necessary rather than sufficient conditions. In addition we shall specify a set of axioms on our preference among the alternatives which are both necessary and sufficient for the expected utility criterion to hold. The relationship to other approaches will then be considered.

For the expected utility criterion to hold one of the assumptions is that the probabilities of the consequences for each choice is known or, at least, the preference among our choices is consistent with the existence of probabilities for which the expected utility may be calculated. Critics of the expected utility criterion have suggested that a knowledge of the probabilities of the consequences occurring in addition to their specific values, ought to influence the preference of the ordering. As is, the expected utility criterion does not hold for alternatives where probabilities are not known. This case, therefore, includes nearly all practical decision situations.

A second difficulty arising from the assumption is that probabilities may not exist for receiving certain rewards for some of the actions. Clearly the expected utility criterion can not then be used. Other alternatives must be used. In Part II of the thesis we shall consider one such alternative.

1.0 Introduction to Part I

In Part I of the thesis we are concerned with conditions that must be made on the ordering so that the expected utility criterion holds.

Various sets of assumptions can be postulated from which an expected utility criterion can be concluded. We shall consider five of the most prominent approaches here, including those of vonNeumann-Morgenstern (1947) and Savage (1954).

The differences among the approaches are due to particular assumptions made about the probability space and about the space on which preference is defined. All these approaches are concerned with sufficient conditions on the ordering rather than necessary conditions. For example, the Savage approach implies that the utility function under consideration must be bounded. This is clearly not a necessary condition.

Throughout the thesis we shall make the following distinction: If an assumption is made with regard to the preference ordering on A, indirectly or directly, the assumption will be called an axiom.

The set of axioms we shall consider can be stated in a simplified form as follows:

> Axiom I. There exists a real-valued function f on the action set A that preserves the ordering on A. That is, if a is preferred to b, then $f(a) > f(b)$, and if we are indifferent between a and b, then $f(a) = f(b)$. We shall discuss this axiom with its implications in detail in section 3.

Axiom II. The function f can be decomposed over the set of events into a sum of functions h, each of which depends only on the pay-off on that particular event. For example, if a results in one of the pay-offs $r_1,\ldots,r_n$ for the events $B_1,\ldots,B_n$ then

$$f(a) = h(r_1,B_1) + \ldots + h(r_n,B_n).$$

We shall discuss this axiom in greater detail in section 4.

Axiom III. The function h can also be decomposed into the product of two functions, W on β and U on R. That is, Axiom III assumes

$$h(r,B) = U(r)\cdot W(B)$$

where $W(B) \geqslant 0$ for all $B \in \beta$.

One of the implications of this axiom is that if a and b are two actions whose payoffs are r and s respectively for all states of nature then

$$f(a) - f(b) = h(r,\Omega) - h(s,\Omega)$$

$$= U(r)W(\Omega) - U(s)W(\Omega)$$

$$= (U(r) - U(s))W(\Omega).$$

Therefore $f(a) > f(b)$ if and only if U(r) is larger than U(s). Hence an ordering may be specified on R such that $f(a) > f(b)$ if and only if r is preferred to s (i.e., if $U(r) > U(s)$). In more general terms,

$$h(r,B) > h(s,B)$$

if and only if r is preferred to s.

Therefore if f(a) is thought of as our evaluation of alternative a on Ω, $h(\cdot,B)$ can be thought of as our evaluation of the alternatives on the set B. Similarly, consider the difference

$$h(r,B) - h(r,C) = U(r)(W(B) - W(C))$$

and assume that the reward r is such that $U(r) > 0$. Then $h(r,B) > h(r,c)$ if and only if $W(B) > W(c)$. Consider, for example, that if r = \$100 and B and C are some arbitrary events, then

h(r,B) = "our evaluation of receiving \$100 if event B occurs"

and

h(r,C) = "our evaluation of receiving \$100 if event C occurs".

Then our evaluation of the first is greater (or preferred) to the second if $W(B) > W(C)$. Clearly this would be the case if the likelihood of B occurring is greater than the likelihood of C occurring. Therefore, it seems necessary to relate W(B) to the probability of B. Axiom IV specifies this relation.

<u>Axiom IV</u>. $W(B) = \mu(B)$.

In section 7 we shall show that these axioms are both necessary and sufficient for the ordering on A to correspond with the ordering induced by the expected utility.

To summarize, the objective in the first part of the thesis is:

1) To specify a necessary set of axioms for the expected utility criterion to hold;
2) To indicate the implications of these axioms in terms of the standard decision problems where the choice by many would contradict the expected utility criterion, and to provide a short summary of empirical studies with regard to the acceptance level of each axiom;
3) To relate, when practical, this set of axioms to those made by others.

2.0 Assumptions and Notations

In section 2.1 we shall summarize the notation terminology and assumptions used throughout Part I of the thesis. The same notations are also used in Part II of the thesis although the assumptions are modified. In section 2.2 we shall specify some of the most common approaches to the expected utility theory.

2.1 Basic notations and assumptions

We shall first assume that there exists a reward set R, whose elements will be denoted by r,s,t,v, etc. R is not necessarily a set of monetary rewards nor are the rewards necessarily desirable. We do, however, assume that R is a well-defined set, i.e., that the elements are definite and distinct.

We also assume the existence of a state space Ω consisting of elements ω, called states. This set represents what may possibly happen in the future, and can not be controlled or influenced. There is, in addition, a reward function $X(\cdot,a)$ from Ω to R, by which we mean reward $X(\omega,a)$ is obtained if state ω occurs and action a is chosen. Let Γ denote all functions from Ω to R being considered, i.e., the elements of Γ are $X(\cdot,a)$, $X(\cdot,b)$, etc. The set $\{a,b,...\}$ is denoted by A and is called an action set; that is, $\Gamma = \{X(\cdot,a):a \in A\}$.

The existence of these sets makes up the basic ingredients in decision theory. There are also some additional basic mathematical assumptions which must be made. We shall summarize these jointly in Assumption 1. (For a definition of the terminology, see Appendix II.)

<u>Assumption 1</u>. There exist:

i) a probability space (Ω,Θ,μ), where Ω is the set of <u>states</u>,

ii) a measurable space (R,Ψ), where R is the set of <u>rewards</u> or outcomes,

iii) an index set A, called an <u>action space,</u> such that for each $a \in A$ there exists a reward function $X(\cdot,a)$ from Ω to R,

iv) a σ-algebra β of Ω such that $\Theta \subset \beta$ and for each $a \in A$, the function $X(\cdot,a)$, is β-measurable, and

v) a relation $\overset{A}{\leqslant}$ on A.

Because of the importance of relations and orderings in expected utility theory, we have summarized the definitions and properties of the more common ones in Appendix I. We shall also use the relation $\overset{A}{<}$, which we define as:

$$a \overset{A}{<} b \text{ if and only if } a \overset{A}{\leqslant} b \text{ and not } b \overset{A}{\leqslant} a.$$

The relations $\overset{A}{<}$ and $\overset{A}{\leqslant}$ are interpreted as preferences among the actions. That is, if action a is preferred to action b, we simply write $b \overset{A}{<} a$ or $b \overset{A}{\leqslant} a$. Preferences exist without knowing exactly what state will occur.

With this assumption, expected utility theory is concerned with the conditions and additional assumptions under which there exists a measure W on β, such that $W(B) = \mu(B)$ if $B \in \Theta$;

and further, given a preference on A (that is a relation $\overset{A}{<}$ on A) under what conditions and assumptions there exists a real valued function U on R such that a $\overset{A}{\leqslant}$ b implies

$$\int U(X(\omega,a))d\mu \leqslant \int U(X(\omega,b))d\mu.$$

Variations of this problem occur when $\Theta = \beta$, or when the ordering is defined on Γ or Π rather than on A, where $\Gamma = \{X(\cdot,a): a \in A\}$ and Π = {all probability measures induced on R, by the functions in Γ}.

The assumptions necessary for the expected utility criterion to hold are of two types. The first specifies the mathematical assumptions. For example, how large can β be, or what actions can be included in A? Those of the second type are far more important in that they specify the decision maker's behaviour. That is, the extent to which his preference in one situation also specifies his preference in another. We are most interested in the latter type of assumption. Both sets of assumptions are, of course, interrelated and cannot in general be separated. However, for each assumption we shall consider our main concentration to be on the implications of the ordering.

2.2 Approaches to the axioms of expected utility theory

Various sets of assumptions or axioms can be postulated from which an expected utility criterion can be concluded. We shall consider five of the most prominent approaches here:

1. vonNeumann-Morgenstern (1947)
2. Marschak (1950)

3. Savage (1954)

4. Arrow (1971)

5. Luce and Krantz (1971)

The axioms are given in Appendix III and we shall merely summarize the basic differences here. These differences are due to particular assumptions being made about the probability space and about the space on which a preference is defined. Each approach represents a different assumption about the space of preference orderings.

vonNeumann-Morgenstern Axioms. The vonNeumann-Morgenstern (1947) approach does not make any assumptions directly on the underlying probability space (Ω,Θ,μ), or on Γ. Instead it assumes that Π, the set of probability measures induced on R by members of Γ, is equal to the set of all discrete probability measures on R and that the probabilities on all states are known, that is, $\Theta = \beta$. It assumes an ordering on R, which is then extended to Π.

Marschak Axioms. Marschak (1950) was the first to adopt an approach which establishes an ordering on the probability measures. Samuelson (1952), Herstein and Milnor (1954) and other authors have also adopted this formulation. The axioms considered in the appendix are essentially the same as Jensen's (1967) axioms, and he has shown them to imply Marschak's axioms.

In this approach, we also ignore the underlying probability space since all assumptions are based on the induced probability measure. Therefore, this approach also assumes that the probabilities of all states of nature are known.

Savage Axioms. Savage (1954) starts with a measurable space (Ω,β). In his approach an ordering is assumed on Γ such that a probability measure can be derived on β. Further, it is assumed that using this probability measure the ordering satisfies the expected utility criterion. In this case, $\Theta = \{\Omega,\phi\}$. That is, probabilities are only known for the universal set and the empty set.

Arrow Axioms. Arrow (1971) basically uses the Savage axioms; however, Arrow specifies the ordering on A rather than on Γ.

Luce and Krantz Axioms. One argument criticizing the Savage Axioms has been that all the random variables have been defined on the same state space. For example, the set Ω of states of the world appropriate for considering betting on heads in a coin flip is quite different from the set Ω appropriate for considering investing in a particular stock. Luce and Krantz (1971) developed an axiomatic system to handle this case by considering the ordering on the set $\Gamma_\beta = \{\Gamma_B | B \varepsilon \beta\}$ where Γ_B contains the functions in Γ with their domains restricted to B. We denote these functions by $X_B(\cdot,a)$, $a \varepsilon A$.

Some general relations exist between the different approaches. For example, Marschak and vonNeumann-Morgenstern both assume that a numerical probability is given, while the others do not make this assumption. However, this is not a fundamental difference since by adding some axioms we can always derive a probability based on preference. This will be discussed in greater detail in section 6.3. The other approaches have already included the axioms needed to derive the probability.

Arrow's and Savage's axioms are nearly identical except for the space in which they are defined, though one readily translates into the other by the real valued function which is assumed to exist between A and Γ. The main difference is that the Savage axioms do not necessarily imply that the probability measure is σ-additive. This will be discussed in section 6.3 and is also considered in Part II.

There is very little difference between the Luce and Krantz axioms and Savage's. First, Savage assumed that all reward functions have the same domain. Second, he assumed that a decision does not affect the probabilities of the states of nature. The insignificance from a theoretical viewpoint of these differences in our framework is illustrated as follows:

Suppose, for example, that a,b ε A are the actions such that their respective reward functions are defined by

$$X(\omega,a) = \begin{cases} 1 & \text{heads occurs when a coin is flipped} \\ 0 & \text{otherwise} \end{cases}$$

and

$$X(\omega,b) = \begin{cases} 1 & \text{if 3 occurs when a die is rolled} \\ 0 & \text{otherwise.} \end{cases}$$

In the Luce and Krantz approach we would only need to consider the state space {heads occurring, heads not occurring} for a, and for b we consider the state space {3 occurring, 3 not occurring}.

In Savage's approach we can not do this. We would have to consider all possible states (heads, 1), (heads, 2),..., (heads, 6) and (tails, 1), (tails, 2),...,(tails, 6) for each action. It is, perhaps, more difficult to evaluate the probabilities in the Savage case since we would have more sets. The difference is only a question of how to derive the subjective probabilities. There would only be a fundamental difference between Savage's and Luce and Krantz's axioms if there existed a set of functions defined on a subset of Ω for which we can not construct a function defined on all of Ω with the same distribution on R. Since this can always be done, the difference can be ignored.

We shall consider other similarities between the specific axioms from the different approaches, as we relate them to the axioms presented here.

3.0 Ordering axiom

In this section we shall consider the first of the four axioms necessary for the expected utility criterion to hold. In section 3.1 we shall state the axiom, in section 3.2 some of the implications of the axiom are studied, in 3.3 its relationship to the five sets of sufficient axioms is examined, and finally in section 3.4 alternative related axioms are explored.

3.1 Statement of Axiom I

In the axioms given in the introduction, we have chosen to define the ordering on the action set A which is the most general set we can choose for this purpose.

<u>Axiom I</u>. Existence axiom

There exists a real-valued function f on A such that for any $a,b \in A$,

$$\text{if } a \overset{A}{\leqslant} b \text{ then } f(a) \leqslant f(b), \text{ and}$$
$$\text{if } a \overset{A}{<} b \text{ then } f(a) < f(b).$$

The function f shall be called the <u>evaluation function</u>.

Several theorems can be found which specify the necessary conditions for the existence of f; see, for example, Debreu (1954) and Peleg (1970). We give a short summary of them in Appendix I. The conditions for f to exist can be divided into two types: (i) conditions on the relation $\overset{A}{<}$ and (ii) topological conditions.

For example, if our preferences are as follows: $b \overset{A}{<} a$ and also $c \overset{A}{<} b$ then by assuming the existence of f, this implies that c can not be preferred to a. That is, the relation on A "inherits" certain properties of the natural ordering of real numbers through f.

For most practical purposes, an individual utility function must be derived by some method. We shall not go into details as to how this is generally done. However, there are some implications on the ordering which we need to consider. In deriving a utility function a set of relatively easily made choices of some alternatives is determined in such a way that a utility function may be specified. This function may not be the same as the utility function (if it exists) when we consider the alternatives in A. This creates difficulties from a practical viewpoint, i.e., an individual utility function may depend on the alternative which is being considered. To avoid this, we shall assume that A always contains those actions needed for deriving a utility function, and secondly that the evaluation function is always the same for a given action. This implies that we evaluate an action on its own merits without considering its relations to other actions. In mathematical terms this is equivalent to considering only the ordering on subsets of A which have been induced by $\overset{A}{\leqslant}$. That is, if $A_o \subset A$ then the only ordering we shall consider on A_o is defined by

$$\overset{A_o}{\leqslant} = \overset{A}{\leqslant} \cap A_o \times A_o$$

or, if $A_1 \supset A$ then we only consider orderings on A_1 such that

$$\overset{A}{\leqslant} = \overset{A_1}{\leqslant} \cap A \times A.$$

Orderings induced in this way we shall call <u>hereditary orderings</u>.

The topological implications are concerned with what is meant by a sequence of actions approaching a given action. This is considered in Appendix I. Another topological property on A is the cardinality of the set of indifference classes. If A_1 is the subset of A such that we are never indifferent between any two of the alternatives in A_1, then the cardinality of A_1 must be less than or equal to the cardinality of the real numbers, otherwise f clearly does not exist. This condition we shall assume always holds in section 4.2. From our point of view we are most interested in whether Axiom I is a reasonable assumption for an individual to make. Therefore we are mainly interested in the implications on the ordering.

3.2 Implications of some assumptions concerning $\overset{A}{\leqslant}$

In this section we shall consider the implications of the ordering from a decision maker's point of view. Hence the question we shall consider is whether it is "reasonable" to expect the decision maker to have a transitive preference or a complete or a partial ordering or, for that matter, if it is acceptable to assume that the ordering is hereditary. We shall consider these issues in turn.

Transitivity. In Appendix I transitivity is defined as a binary relation such that if the property holds between a and b and between b and c, it must also hold between a and c. Rather than discuss whether this is a reasonable or unreasonable assumption to make we shall consider a few examples where transitivity holds and a few where it does not. We shall start with some trivial examples, and then discuss some related to some decision problems. In doing so, we hope for a better understanding of the implications of Axiom I. Note that Axiom I does not quite imply transitivity; that is, if there exist preferences between a and b, b and c and also between a and c, the preferences must be transitive.

Some of the most common examples of transitive relations are the physical properties such as 'heavier than', 'smaller than', 'longer than' and so on. An example of a non-transitive relation is 'father of'. Less trivial examples of the non-transitivity in decision theory can be categorized into either of the following two cases. The first case of non-transitive ordering occurs when the reward set R is formed by a Cartesian product and the ordering is induced on R by an ordering on each component in the Cartesian product separately. The second case of non-transitive ordering occurs when the ordering is induced by a probability measure. We shall illustrate both these approaches below.

To consider the first case, suppose that the reward set is equal to the product of three sets, R x S x T. The reward for action a can then be written as the triple (r_a, s_a, t_a)

where $r_a \in R$, $s_a \in S$ and $t_a \in T$. If individual orderings are first defined separately on R, S and T, the following preference may occur for actions a, b and c:

$r_a \overset{R}{<} r_b$	$r_b \overset{R}{<} r_c$	$r_a \overset{R}{<} r_c$
$s_a \overset{S}{<} s_b$	$s_c \overset{S}{<} s_b$	$s_c \overset{S}{<} s_a$
$t_b \overset{T}{<} t_a$	$t_b \overset{T}{<} t_c$	$t_c \overset{T}{<} t_a$

Hence action b is preferred in two of three consequences over a, similarly c over b, and a over c. Therefore if an overall ordering is defined on the actions by assuming that an arbitrary action c is preferred to another if the first action is preferred on at least two of three rewards, then the induced ordering on the overall reward set is not transitive as our example illustrated since b is preferred to a, c to b but a is preferred to c.

An illustration of this type of intransitivity is as follows: Let our actions be a choice between apple pie, blueberry pie and cherry pie, and let the selection be based on taste, freshness and size. It is easy to see how this may give the contradiction by letting r_a stand for the taste of the apple pie, s_a for its freshness, and t_a for its size; similarly define r_b, s_b and t_b for the blueberry pie and r_c, s_c and t_c for the cherry pie and use the ordering given above.

A variation of this would be to consider three teams, A, B, and C, each team having three players which must play against each player on another team. The players are ranked from 1 to 9 (9 is the best and 1 the worst). Hence if the players on the teams A, B and C are ranked (8,1,6), (3,5,7) and (4,9,2) respectively and if each player meets every other player on the opposing team, we note that A defeats B (winning 5 games out of 9), B defeats C but C defeats A.

A second class of ordering that contradicts transitivity occurs when the ordering is based on probabilities. As an example consider the ordering specified by the relation $a \overset{A}{>} b$ if and only if $\mu\{X(\cdot,a) > X(\cdot,b)\} > 1/2$. The question of transitivity arising here is the following: If $\mu(X > Y) > 1/2$ and $\mu(Y > Z) > 1/2$, is it also true that $\mu(X > Z) > 1/2$? The answer is no as the following illustration shows: Let

$$X(\omega,a) = 1 \quad \omega \in \Omega$$

$$Y(\omega,b) = \begin{cases} 5 & \omega \in B \\ 0 & \omega \in \overline{B} \end{cases}$$

$$Z(\omega,b) = \begin{cases} -1 & \omega \in D \\ 2 & \omega \in B \\ 4 & \omega \in \overline{B}-D \end{cases}$$

where $D \subset \overline{B}$. If we assume that $\mu(B) = 1/2 - \varepsilon$ and $\mu(D) = 2\varepsilon$, where $0 < \varepsilon < 1/4$, then $\mu(X > Y) = \mu(\overline{B}) = 1/2 + \varepsilon$, and $\mu(Y > Z) = \mu(B) + \mu(D) = 1/2 + \varepsilon$, but $\mu(X > Z) = 2\varepsilon < 1/2$. Hence $b \overset{A}{<} a$ and $c \overset{A}{<} b$ but $a \overset{A}{<} c$.

Another illustration is the following game. Two of the four dice shown in Fig. 3 are rolled and the die with the higher number wins (Gardner, 1974).

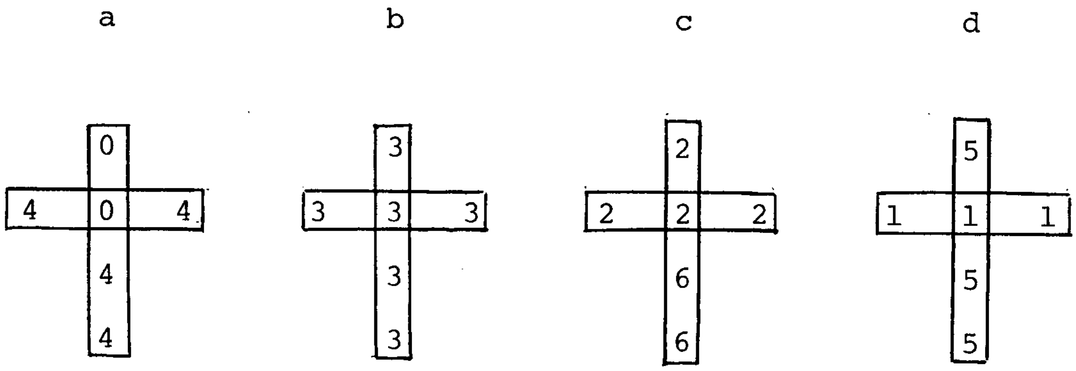


Fig. 3.1

A non-transitive set of dice

Note that die a will win over die b with the probability 2/3; and similarly b wins over c, and c wins over d. However, d wins over a with the probability 2/3.

For an interesting study of transitivity, see May (1954). He considered several examples of non-transitive relations, e.g., the case of a pilot faced with the choice of flames or red-hot metal, red-hot metal or falling, or falling or flames.

In the examples above there exists no order preserving, real-valued function on the set A of alternatives. There is a strong argument for accepting transitivity which is described in Raiffa (1961). The existence of the evaluation-function is a means by which we can determine a value for each alternative. In most practical cases we assume this is a function of some monetary value, i.e., we would be willing to pay more for the choice of a particular given alternative rather than another one. Raiffa used this argument to support the proposition that transitivity is a reasonable assumption. For example, suppose preferences are non-transitive, i.e., a is preferred to b

b is preferred to c

and c is preferred to a,

Raiffa then argues that if the individual is given alternative a, he would be willing to pay a small amount to switch to c since c is preferred to a. Once c is obtained, Raiffa again assumes for the same reason that one would be willing to pay a small amount to switch to b, from b to c, and so on, with a small amount being paid for each switch. Since it is obvious that one would not be willing to continue paying a small amount for each switch, Raiffa concludes that transitivity must be satisfied by the ordering. However, assuming that all alternatives can be compared on a monetary scale is tantamount to assuming the existence of an evaluation function. Therefore Raiffa assumes the ordering is transitive if a real-valued function exists.

Partial ordering vs. complete ordering. One theorem in Appendix I which used a partial ordering rather than a complete ordering is due to Peleg (1970). However, the idea behind a partial ordering is due to Aumann (1962) who first proved an expected utility theorem without using a complete ordering, but since Peleg's theorem is a little more general we quoted his. The question Aumann raised was: "Does rationality demand that an individual make definite preference comparisons between all possible actions?" As an example he gives his preferences: "I prefer a cup of cocoa to a 75-25 lottery of coffee and tea, but reverse my preference if the ratio is 25-75." However, can a break-even point be determined between a lottery and the cocoa?

For our purpose this is irrelevant since with the assumption of the existence of f, a complete ordering may be induced on A which corresponds to any partial ordering for those elements which can be compared. Therefore, the conditions on f are identical in both cases.

Hereditary ordering. With all hereditary orderings we assumed that an alternative may be removed without affecting the relative ordering of the remaining alternatives. If, for example, A = {a,b,c} and a is preferred to b, then if alternative c is not possible, i.e., A = {a,b} then a ought to be still preferred to b. In some criteria in decision making under uncertainty this is not the case.

Consider, for example, Savage's regret criterion in the following example. We are given a choice between actions a, b and c with the payoffs as in the following payoff matrix.

	Event	
Action Space	B	$\overline{B}$
a	3	6
b	0	10
c	6	0

The regret matrix is formed by subtracting each entry in each column from the largest entry in the column and hence would give the matrix:

	Event	
Action Space	B	$\overline{B}$
a	3	4
b	6	0
c	0	10

The maximum regret for actions a, b and c would be 4, 6 and 10 respectively and we order the actions by choosing the minimum of the maximum regrets, hence the ordering becomes $c \overset{A}{<} b \overset{A}{<} a$. However if action c is not included the regret matrix would change such that our ordering would be $a \overset{A}{<} b$.

Another example has been suggested by Luce and Raiffa (1957) which contradicts the hereditary ordering.

"A gentleman wandering in a strange city at dinner time chances upon a modest restaurant which he enters uncertainly. The waiter informs him that there is no menu, but that this evening he may have either broiled salmon at $2.50 or steak at $4.00. In a first-rate restaurant his choice would have been steak, but considering his unknown surroundings and the different prices he selects the salmon. Soon after the waiter returns from the kitchen, apologizes profusely, blaming the uncommunicative chef for omitting to tell him that fried snails and frog's legs are also on the bill of fare at $4.50 each. It so happens that our hero detests them both and would always select salmon in preference to either, yet his response is, "Splendid, I'll change my order to steak."

A justification for the gentleman changing his order can be given as follows. He enters an unknown restaurant without knowing the quality of food he expected to be served. He played it safe and decided bad salmon is better than steak for the price specified. Once he found out that the restaurant also has frogs' legs (i.e., is a better class restaurant), he decided this restaurant would not serve a bad steak or a bad salmon. Hence his choice becomes a good steak or a good salmon for the specified prices, and therefore his choice changes. Hence, he obtains more information by the addition of the new action. In general when we speak of hereditary ordering we assume no additional information is given by adding another action.

Another common example stated is the voting paradox. Let 1/3 of all voters have the preference of the three candidates a, b and c as $c < b < a$, another third of the voters the preference $a < c < b$, and the final third $b < a < c$. Hence 2/3 of the voters prefer b to c, and 2/3 prefer c to a, and 2/3 prefer a to b. This interesting result is perhaps in relation to the hereditary assumption. If a did not run, b would win; if b did not run, c would win; and if c did not run, a would win.

3.3 Ordering properties in the different approaches

We have stated how each of the different approaches we consider here implies the expected utility theory but uses different axiom systems to prove an expected utility theorem. Here we shall compare the different approaches to Axiom I.

We first note that:

1) vonNeumann-Morgenstern define preference orderings on the set R,
2) Marschak defines preference orderings on the set of all discrete probability measures on R,
3) Savage defines preference orderings on the set of all possible functions from Ω to R,
4) Arrow defines preference orderings on the set of actions A,
5) Luce and Krantz define preference orderings on an arbitrary set of functions Γ_β.

Each of these axiom systems assumes a complete ordering.

The most general of these assumptions is Arrow's. For example, a and b may be two actions belonging to A such that $X(\omega,a) \equiv X(\omega,b)$ for all $\omega \in \Omega$ where $f(a) \neq f(b)$ but clearly $X(\cdot,a) = X(\cdot,b)$. Similarly two functions $X(\cdot,a)$ and $X(\cdot,b)$ may have the property $X(\omega,a) \neq X(\omega,b)$ for all $\omega \in \Omega$ but $P(C,a) = P(C,b)$ for all $C \in \Psi$, i.e., the functions $X(\cdot,a)$ and $X(\cdot,b)$ induce the same probability measure on the rewards but $X(\cdot,a) \neq X(\cdot,b)$. Therefore, some decision makers would find it easier to accept an ordering on the reward set R rather that on the action set A, for example. The reason for this is obvious: the cardinality of the set for the most general cases are as follows:

$$C(R) \leqslant C(\Pi) \leqslant C(\Gamma) \leqslant C(\Gamma_\beta) \leqslant C(A),$$

where C(R) denotes the cardinality of R, and similarly for

$C(\Pi)$, $C(\Gamma)$, and so on. That is, if we assume an ordering on A, we can induce an ordering on Γ with a subset of A, and similarly, if we assume an ordering on a probability distribution on R, we can induce an ordering on R with a subset of Π. Therefore, the general assumption of an ordering on A is a stronger assumption than an ordering on R.

There also exist additional assumptions on the cardinality. For example, vonNeumann-Morgenstern assume that between every two rewards there exists another, i.e., if $r_1, r_2 \varepsilon R$, $r_1 \overset{R}{<} r_2$, there exist $r_o \varepsilon R$ such that $r_1 \overset{R}{<} r_o \overset{R}{<} r_2$. Marschak assumes that all probability measures on R belong to Π. Savage assumes that all functions from Ω to R belong to Γ. All these assumptions are not necessary. It is sufficient to assume that the function f exists on A, and the ordering may be represented by the numerical value of the function.

3.4 Alternatives to Axiom I

Most normative theories of decision-making accept Axiom I; that is, the existence of a real-valued evaluation function. Minimax, maximax and expected values, for example, all satisfy Axiom I. Similarly many of the criteria in finance such as the pay-back method, the net-present value, and the internal rate of return all assume Axiom I. As a matter of fact, very few alternatives to Axiom I can be presented. Two different approaches are discussed below.

Stochastic utility theory. Some empirical studies have shown that many people are not consistent in repeated choice situations. That is, sometimes they prefer a to b and sometimes the opposite. This has led to what is called the "stochastic utility theory". In this theory, the axioms are stated in terms of probabilities of choice. For instance, Debreu's axioms (1958) are as follows, where a, b, c and d are arbitrary actions:

<u>Axiom 1</u>: S is a set; p is a function from S x S to (0,1) such that $p(a,b) + p(b,a) = 1$.

<u>Axiom 2</u>: $(p(a,b) < p(c,d))$ implies $(p(a,c) < p(b,d))$.

<u>Axiom 3</u>: If $p(b,a) < q < p(c,a)$ where q is any real number, then there is $a,d \in S$ such that $p(d,a) = q$.

The interpretation of p(a,b) is that a is preferred to b, a proportion p(a,b) of the time. Debreu proves that these axioms imply that there exists a real-valued function V on S such that

$$p(a,b) < p(c,d) \text{ if and only if } V(a) - V(b) < V(c) - V(d).$$

Clearly the function V can not be considered as an evaluation function of actions since this implies that if $V(a) > V(b)$ then $p(a,b) = 1$, and hence p must be a function to $\{0,1\}$.

From our discussion on ordering induced by probability, it follows that if $p(a,b) > 1/2$ and $p(b,c) > 1/2$ it is not necessary that $p(a,c) > 1/2$, i.e., if action a is preferred more often to action b, and if b is preferred more often to action c, it does not follow that a is preferred more often to action c.

<u>Multivariant evaluation functions</u>. Another alternative to Axiom I is to assume f as a function from A to E^n (the n-dimension Euclidian space). In mapping into E^n more variations can be defined on the ordering than on E as we do not need to satisfy transitivity. Consider, for example, the relation < on E^2 defined by $(\alpha,\beta) < (\gamma,\delta)$ if and only if $\beta^2\gamma^2 < \alpha^2\delta^2$. For this relation a real-valued function satisfying Axiom I does not exist.

As an example to show that n-dimension Euclidian space is sometimes appropriate, one might consider a group of individuals where each one satisfies Axiom I on the action set. An evaluation function f can then be defined as $f(a) = (f_1(a), f_2(a), \ldots, f_n(a))$ where $f_i(a)$ is the i^{th} individual's evaluation function. Sometimes this function is then reduced to E^1 by

1) $a \overset{A}{<} b$ if $\sum_{i=1}^{n} f_i(a) < \sum_{i=1}^{n} f_i(b)$.

(Hence this is reduced to Axiom I and is used in finance for net-present value.)

2) $a \overset{A}{<} b$ if $f(a)$ L $f(b)$

where L is the relation "lexiographically larger". This approach is used in social choices.

4.0 Additivity Axiom

In this section we shall discuss Axiom II, starting by stating the axiom in section 4.1. The axiom has some very strong implications and in section 4.2 we shall consider some "paradoxes" that have been proposed in relation to it. In section 4.3 we shall summarize some empirical studies associated with some of these paradoxes, and in section 4.4 we consider the relationship to other axiom systems. Finally, in section 4.5 some alternatives to Axiom II are stated.

4.1 Statement of Axiom II

Let us assume that a given action $a \varepsilon A$ results in one of the following possible rewards: $r_1,\ldots,r_n$ if the event $B_1,\ldots,B_n$ occurs respectively. Axiom II as stated in the introduction asserts the existence of a real-valued function h, satisfying the identity

$$f(a) = h(r_1,B_1) +\ldots+ h(r_n,B_n).$$

As a generalization of this axiom, we shall allow for the reward function $X(\cdot,a)$ to result in any of an uncountable number of rewards on any set $B \varepsilon \beta$. Note that the ordering we defined on A could have been defined on Γ, if there is a one-to-one function between A and Γ, i.e., we could have written $f(X(\cdot,a))$ for $f(a)$. The reason for not doing so was both to emphasize the fact that we chose an action a, which gave us the reward function $X(\cdot,a)$ and also the simplicity of notation. However, a further understanding can be achieved by considering the axioms in terms of random variables. Consider the set Γ_B, i.e., the set of all functions $X_B(\cdot,a)$, $a \varepsilon A$ where $X_B(\cdot,a)$ is the restriction of $X(\cdot,a)$ to B. Axiom I assumes that a relation exists on Γ, or at least that an ordering may be induced. A natural extension of this assumption would be to assume that there also exists an ordering on Γ_B for any $B \varepsilon \beta$, as Luce and Krantz did, and a real-valued order preserving function on Γ_B. Axiom II makes this assumption in terms of actions rather than reward functions.

Axiom II therefore assumes first the existence of a uniquely defined real-valued function $h(B,a)$ on $\beta \times A$, where $h(B,a)$ may be regarded as the evaluation function of the reward function $X_B(\cdot,a)$. It also specifies the relation between $f(a)$ and $h(B,a)$. If, for example, $a,b \varepsilon A$, and $h(B,a) = h(B,b)$, that is, we are indifferent between action a and action b if event B occurs, and if $h(\overline{B},a) > h(\overline{B},b)$, that is we prefer action a to action b if event $\overline{B}$ occurs, Axiom II concludes that action a is preferred to action b.

It is clearly not necessary to assume that an ordering exists on Γ_B. However, since $h(B,a)$ is assumed to exist, it is perhaps easiest to consider $h(B,\cdot)$ as the evaluation function of an ordering on Γ_B.

Axiom II. Additivity axiom

There exists a real valued function h on $\beta \times A$ such that

a) $f(a) = h(\Omega, a)$

b) for $\{B_i\}$ $i=1,\ldots$ such that $B_i \varepsilon \beta$ for all i and $B_i \cap B_j = \emptyset$ for $i \neq j$, then $h(\bigcup B_i, a) = \sum_{i=1}^{\infty} h(B_i, a)$ for all $a \varepsilon A$

c) for any $B \varepsilon \beta$, and for any $a,b \varepsilon A$ such that $X_B(\cdot,a) = X_B(\cdot,b)$, then $h(B,a) = h(B,b)$.

The necessity of part b) of this axiom follows from the fact that the integral is σ-additive, that is, if we assume $\{B_i\}$ is a partition of Ω such that $B_i \varepsilon \Theta$ and if we assume for the time being (as we shall prove in later sections) that

$$h(B_i, a) = \int_{B_i} UX(\omega, a)\,d\mu,$$

then clearly if f(a) is equal to the expected utility of $X(\cdot,a)$ we have the following identities:

$$\begin{aligned} f(a) &= EUX(\cdot,a) = \int UX(\omega,a)\,d\mu \\ &= \sum_{i=1}^{\infty} \int_{B_i} UX(\omega,a)\,d\mu \\ &= \sum_{i=1}^{\infty} h(B_i,a). \end{aligned}$$

Part c of the axiom specifies in terms of integral that if two functions are equal they must have the same integral value.

In the normal approach to integration $h(B,a)$ is defined for functions which are constant on B. An extension is then made to functions which only take finitely many values. Finally, extensions are made to functions which take uncountably many values. Here we have chosen to reverse this approach since we have assumed $f(a)$ already exists before we evaluate it.

Some comments on Axiom II. Several questions arise from Axiom II of a rather technical nature in regard to $\sum h(B_i,a)$. The reason for this is that h is not necessarily a non-negative function and it may, therefore, be important in what order the B_i's are selected. For example, assume that $B_1, B_2, \ldots$ is a partition of B such that the sequence $h(B_1,a) + h(B_2,a) + h(B_3,a) \ldots$ is equal to

$$1 - 1/2 + 1/3 - 1/4 + \ldots .$$

Then by changing the order of the B_i's, we could have the sequence

$$1 + 1/3 - 1/2 + 1/5 + 1/7 - 1/4 + \ldots$$

and although both sequences converge, they do not converge to the same value.

Therefore by assuming that

$$h(B_i,a) = \sum h(B_i,a)$$

we assume that the summation is a constant value not only for any partition of B, but also for any rearrangement of a given partition. This assumption is equivalent to assuming that the sequence $\sum_{i=1}^{n} h(B_i,a)$ converges absolutely (see, for example, W. Rudin (1969), pp.68-69):

<u>Theorem</u>. a) If $\sum a_n$ converges for all rearrangements, then they all converge to the same sum.

b) $\sum a_n$ converges for all rearrangements if and only if $\sum a_n$ converges absolutely.

4.2 Implications of Axiom II

There are several important implications of this axiom. For example, it implies the existence of a utility function, if some regularity conditions are assumed. Other implications are of equal importance in that they specify very strong conditions on the preference ordering on A. We shall first specify these mathematical implications and secondly illustrate the "paradoxes" which contradict the axiom.

<u>Lemma 4.2.1</u>. $h(\emptyset,a) = 0$

<u>Proof</u>. Let $B_1 = \emptyset$, B_i, i=2,3,... be a sequence of sets such that $B_i \cap B_j = \emptyset$ $i \neq j$, then

$$h(\bigcup_{i=1}^{\infty} B_i,a) = h(\emptyset,a) + \sum_{i=2}^{\infty} h(B_i,a)$$

and since $B = \bigcup_{i=2}^{\infty} B_i = \bigcup_{i=1}^{\infty} B_i$

$$h(\bigcup B_i,a) = \sum_{i=2}^{\infty} h(B_i,a)$$

hence since $\left|h\left(\bigcup_{i=1}^{\infty} B_i, a\right)\right| < \infty$ then

$$h(\phi, a) = 0.$$

Lemma 4.2.2. $h(\cdot, a)$ is finitely additive.

Proof. This follows directly since $h(\phi, a) = 0$.

Lemma 4.2.3. h is continuous from below.

If $D_i \subset D_{i+1}$ and $D_i \in \beta$ for $i=1,2,\ldots$ then

$$h\left(\bigcup_{i=1}^{\infty} D_i, a\right) = \lim_{i\to\infty} h(D_i, a).$$

Proof. (This follows from standard measure theoretical results.)

Define $B_i = \left(D_i - \bigcup_{j=1}^{i-1} D_i\right)$ where $B_1 = D_1$.

Hence $\bigcup_{i=1}^{n} B_i = \bigcup_{i=1}^{n} D_i$ for all n, and also $B_i \cap B_j = \emptyset$, $i \neq j$.

Therefore $h\left(\bigcup_{i=1}^{\infty} D_i, a\right) = h\left(\bigcup_{i=1}^{\infty} B_i, a\right) = \sum_{i=1}^{\infty} h(B_i, a)$.

Since $h\left(\bigcup_{i=1}^{n} D_i, a\right) = h\left(\bigcup_{i=1}^{n} B_i, a\right) = \sum_{i=1}^{n} h(B_i, a)$,

we have $\lim_{n\to\infty} h\left(\bigcup_{i=1}^{n} D_i, a\right) = \lim_{n\to\infty} \sum_{i=1}^{n} h(B_i, a) =$

$$= \sum_{i=1}^{\infty} h(B_i, a) = h\left(\bigcup_{i=1}^{\infty} B_i, a\right) =$$

$$= h\left(\bigcup_{i=1}^{\infty} D_i, a\right).$$

The properties we have discussed so far are based on the second part of the axiom. There is also one property based on the third part which is important to the theory.

To prove this, however, an additional assumption is needed, that is, that all constant functions belong to Γ.

Assumption 2. If $Y(\omega) \equiv r$ for all $\omega \in \Omega$ and for any $r \in R$, then there exists an $a \in A$ such that $X(\omega,a) = Y(\omega)$.

This assumption is not necessary for the expected utility criterion to hold and we shall not assume that it holds in general. One implication of this assumption is that there exists a real-valued function U on R, as we shall show in Lemma 4.2.4. If assumption 2 is not made, however, we must assume that the function U exists on R, which we do in axiom III.

Lemma 4.2.4 If assumption 2 holds, the evaluation function f induces a utility function U on R.

Proof. Assume $X(\omega,a) \equiv r$ and $X(\omega,b) \equiv r$ for all $\omega \in \Omega$ and $a,b \in A$. Part c in Axiom II implies $h(\Omega,a) = h(\Omega,b)$ or $f(a) = f(b)$. Therefore, the function $U(r) = f(a)$ is a uniquely defined function on R.

Although many other results can be proved these are sufficient for our development here.

<u>Implications from a decision maker's viewpoint.</u> Lemma 4.2.1 implies that the number of subsets of Ω that are considered when an action is evaluated is immaterial, that is, we ought to obtain the same value of f(a) independently of what events are considered as long as they jointly include all possibilities. An example of a criterion which does not satisfy this assumption is the Principle of Insufficient Reason. The criterion was first formulated by La Place approximately two hundred years ago, and can be paraphrased as:

> "If no evidence exists that one of the events in a partition is more likely to occur than the others, then the events should be considered equally likely to occur."

This principle is not always accepted because of apparent contradictions such as the following: suppose we flip a coin twice, then four states can occur (H,H), (H,T), (T,H), or (T,T), hence the probability of (H,H) must be 1/4 by the Principle of Insufficient Reason. On the other hand, we can divide the sample space into (H,H) (not (H,H)), and then the probability of (H,H) must be 1/2 or we must have evidence that these events are not equally likely.

Of course, most decision makers would accept the probability of (H,H) being 1/4 as in this case all states can be listed and there is no reason to assume one is more likely to occur than another. In more general situations the question arises as to which partition ought to be made.

Another implication arises from the additivity assumption of h which is related to what is called the Sure-Thing

Principle by Savage (1954) or what is called the Strong Independence Axiom by Samuelson (1952). Several paradoxes are based on the additivity assumption and stem from the following observation: Let a and b be two actions such that $X_B(\cdot,a) \doteq X_B(\cdot,b)$ for some $B \varepsilon \beta$. Since $(B,\bar{B})$ form a partition of Ω we have from Axiom II that

$$f(a) = h(B,a) + h(\bar{B},a)$$

$$f(b) = h(B,b) + h(\bar{B},b),$$

and also $h(B,a) = h(B,b)$. Thus, $f(a) > f(b)$ if and only if $h(\bar{B},a) > h(\bar{B},b)$, that is when two actions are compared we need only evaluate them on the sets on which they obtain different rewards. The decision problem which gives rise to a contradiction in the alternatives is then designed so that most people make a choice such that $f(a) > f(b)$ but $h(B,b) +$ a constant $> h(B,a) +$ a constant. The paradoxes that best illustrate this point are Ellsberg's Paradox II, Allais' paradox, and MacCrimmon's paradox, each of which is described below.

Ellsberg's Paradox II. D. Ellsberg (1961) described the following decision problem. Consider an urn containing 90 balls, of which 30 are known to be red, and the remaining 60 are an unknown mixture of black and yellow balls. One ball is to be drawn at random from the urn, and we are asked to state our preference between a and b, and also between c and d where a, b, c and d are defined as follows:

a: Receive \$1,000 if a red ball is drawn
   Receive \$0 otherwise

b: Receive \$1,000 if a black ball is drawn
Receive \$0 otherwise

c: Receive \$1,000 if a red or yellow ball is drawn
Receive \$0 otherwise

d: Receive \$1,000 if a black or yellow ball is drawn
Receive \$0 otherwise

The paradox can most easily be considered by illustrating the problem in the form of a decision matrix.

	30 Red balls	60 Balls Black balls	 Yellow balls
a	1,000	0	0
b	0	1,000	0
c	1,000	0	1,000
d	0	1,000	1,000

Note that the last event "yellow balls" does not discriminate between the alternatives a and b nor between the alternatives c and d and hence can be ignored. Ignoring this event, we know that a is then identical to c and b is identical to d; hence a choice of a over b would require a choice of c over d. This intuitive argument is the basis of Axiom II which implies its conclusion. In terms of Axiom II, let us denote the event that a red ball is drawn by R, the event that a black ball is drawn by B, and the event that a yellow ball is drawn by Y. Therefore we can express f by Axiom II, part a, as:

$$f(a) = h(R,a) + h(B,a) + h(Y,a)$$

$$f(b) = h(R,b) + h(B,b) + h(Y,b)$$

$$f(c) = h(R,c) + h(B,c) + h(Y,c)$$

$$f(d) = h(R,d) + h(B,d) + h(Y,d)$$

By Axiom II, part b, the following identities must hold:

$$h(R,a) + h(B,a) = h(R,c) + h(B,c)$$
$$h(R,b) + h(B,b) = h(R,d) + h(B,d)$$
$$h(Y,a) = h(Y,b)$$
$$h(Y,c) = h(Y,d)$$

Therefore if $f(a) > f(b)$ this implies

$$h(B_1,a) + h(B_2,a) > h(B_1,b) + h(B_2,b)$$

therefore

$$h(B_1,c) + h(B_2,c) > h(B_1,d) + h(B_2,d)$$

and hence

$$h(B_1,c) + h(B_2,c) + h(B_3,c) > h(B_1,d) + h(B_2,d) + h(B_3,d)$$

or equivalently

$$f(c) > f(d).$$

Of course an even stronger result is implied since $f(a) - f(b) = f(c) - f(d)$, rather than just the inequalities. This implies that if we change the amount we receive if a black ball is drawn until we are indifferent between b and c, then the amount we receive if a black ball is drawn from alternative d must be changed exactly the same to make us indifferent between alternatives c and d. The paradox arises if someone chooses a over b and then d over c. Ellsberg (1961) suggested that this is a likely preference ordering since a known probability of winning is preferred to an unknown change if there is no reason to believe the unknown change has a higher probability of winning.

<u>Allais Paradox</u>. This problem was first proposed by Allais (1953). Again we have to choose between a and b and also between c and d, where

a: \$1 million with a probability of 1.00

b: \$5 million with a probability of 0.10
\$1 million with a probability of 0.89
\$0 with a probability of 0.01

c: \$1 million with a probability of 0.11
\$0 with a probability of 0.89

d: \$5 million with a probability of 0.10
\$0 with a probability of 0.90.

So far in our assumptions we have not considered probabilities but only events. Therefore, we shall rewrite the Allais problem in the form of 100 lottery tickets numbers 1-100 with prizes as given in the payoff matrix below:

	Ticket #1	Tickets #2-11	Tickets #12-100
a	$1 million	$1 million	$1 million
b	0	$5 million	$1 million
c	$1 million	$1 million	0
d	0	$5 million	0

The event of drawing a ticket with any of the numbers 12-100 therefore does not discriminate between a and b nor between c and d.

Therefore by the same arguments used with the Ellsberg paradox we have the relation

$$f(a) - f(b) = f(c) - f(d).$$

To be consistent with Axiom II, a preference of a to b, it must therefore imply a preference of c to d. Of course, it may be argued that since we modified the problem it is not necessarily identical to the original one. This is true since the assumptions we have made do not imply that they are equal. We shall discuss this point in section 4.6.

Allais (1953) suggested that when one alternative gives a certainty (or near certainty) of obtaining a very desirable consequence, one should select it even if it entails passing up a larger amount having a lower probability. When, however, the chances of winning are small and close together, one should take the option that provides the larger payoff. Hence in the above alternative he assumes that we ought to choose a to b but d to c, which contradicts Axiom II.

<u>MacCrimmon's Paradox I</u>. Although both the Ellsberg and Allais paradoxes can be used to illustrate some of the difficulties of Axiom II, the Ellsberg paradox also illustrates the difficulties of assigning a subjective probability to the event B, defined as {a black ball is drawn}. In other words, we can argue that the event $B \varepsilon \beta - \Theta$. We shall study this aspect of the Ellsberg paradox in more detail in section 6.2 and in the second part of the thesis. MacCrimmon (MacCrimmon and Larsson, 1975) has designed a paradox which attempts to capture both Allais' and Ellsberg's arguments into one problem. That is, it has the property of unknown versus known probabilities and also nearly sure as well as very small chances of winning.

Consider an urn containing 100 balls, 20 of which are red; the other 80 are either black or yellow. The number of black balls is between 1 and 5 inclusive. One ball will be drawn from the urn and its colour will determine the payoff received.

Again, choices are to be made between alternatives a and b, and between alternatives c and d.

a: \$100,000 if a red ball is drawn
\$1,000,000 if a yellow ball is drawn
\$0 if a black ball is drawn

b: \$0 if a red ball is drawn
\$1,000,000 if a yellow ball is drawn
\$1,000,000 if a black ball is drawn

c: $100,000 if a red ball is drawn
$0 if a yellow ball is drawn
$0 if a black ball is drawn

d: $0 if a red ball is drawn
$0 if a yellow ball is drawn
$1,000,000 if a black ball is drawn

The problems can be set up in a payoff matrix as follows:

		80 Black or Yellow Balls	
	20 Red balls	1-5 Black balls	75-79 Yellow balls
a	$100,000	0	$1,000,000
b	0	$1,000,000	$1,000,000
c	$100,000	0	0
d	0	$1,000,000	0

Again, we can show that

$$f(a) - f(b) = f(c) - f(d).$$

Therefore the preference $a \overset{A}{>} b$ must also imply $c \overset{A}{>} d$.

4.3 Empirical studies of Axiom II

Most of the empirical studies related to Axiom II have been made in the form of the paradoxes stated in section 4.2. It is not intended to make an extensive survey of these studies but rather to give some indication as to how readily Axiom II has been accepted. Ellsberg's paradox also contradicts Axiom IV, and therefore we postpone the discussion to section 6.3.

Empirical results of models based on the Allais paradox

The particular probabilities and payoff in the Allais paradox have been carefully designed in both monetary values and probabilities to elicit violations of the utility axiom. The model based on the Allais paradox can be represented by the payoff matrix:

	Event		
Action	B_1	B_2	B_3
a:	s	u	t
b:	r	v	t
c:	s	u	w
d:	r	v	w

We showed in section 4.2 that f(a) > f(b) implies f(c) > f(d).

Studies of the common consequence, that is, the reward is the same on a subset of Ω for both actions, have been made by MacCrimmon (1965), Moskowitz (1974), Slovic and Tversky (1975) and MacCrimmon and Larsson (1975). Most of these authors use the Allais problem by having two pairs of choices with $s = u = t$, $r = w$, and with the probability values $\mu(B_1) = 0.01$, $\mu(B_2) = 0.10$, and $\mu(B_3) = 0.89$. MacCrimmon and Larsson varied the probabilities to determine if this would influence the decision maker.

MacCrimmon (1965) study. MacCrimmon uses three common consequence problems. The rewards of all these problems are specified in terms of the occurrence of events rather than in terms of probabilities. In two of the problems, the events are based on standard urns, and from the events most people infer the same probabilities. In the other problem, the occurrence of events was described only verbally (e.g., "very unlikely" or "likely"); the subjects were senior business executives; the problems appeared in the context of investments for a hypothetical business; and the rewards were percent return on the investment. The parameters of the problems may be summarized as:

<u>Problem 1</u>. v = 500% returns on the capital,
$s = u = t$ = 5% returns on the capital,
$r = w$ = bankruptcy.

<u>Problem 2</u>. Same as Problem 1 except for the probabilities which were only qualitative.

<u>Problem 3</u>. v = 75% returns on the capital,
$s = u = t$ = 35% returns on the capital,
$r = w$ = 0% returns on the capital.

In all three problems a significant number of the 36 subjects chose an answer contradicting Axiom II (14, 15, and 13 respectively), and hence confirmed the claim made by Allais. Only nine subjects conformed to the expected utility axioms in all three problems (15 had one deviation, nine had two, and three had three). In addition to their own choices, the subjects were presented with responses supposedly from subjects in a previous session and were asked to criticize them. These results are even more compelling since 29 of the subjects (on problem 1) agreed with the Allais-type answer; only seven agreed with the Axiom II-type answer. However, the hypothesis that in criticizing these responses, subjects tended to pick the answer that most closely corresponded to their own answer, accounts for 32 of the cases.

Moskowitz (1974) study. Moskowitz presented each of 134 students with three problems where the rewards set was defined as the set of possible grades in a course. For the first two problems, Moskowitz does not specify the numerical grades he used in the problems; in the third problem the grades were letter grades. The parameters of his problem three were: v = A grade, s = u = t = B+ grade, r = w = F grade, $\mu(B_1) = 0.01$, $\mu(B_2) = 0.10$, $\mu(B_3) = 0.89$. Moskowitz presented these problems in three different formats: word, tree, and matrix, as suggested in MacCrimmon (1967). He allowed some subjects to discuss the problems in a group while others had to proceed individually. In addition to presenting the problem, he presented pro and con arguments and afterward had the subjects choose again.

Overall, Moskowitz found a rate of violation of about 30%. The tree representation was the most difficult with a violation rate ranging from 20% to 50% (across the other conditions). In the word format, the violation rate was 17% to 40% and in the matrix format the violation rate was 21% to 42%. There were only slight differences in problem types with problem 1 having a violation range of 17% to 46%; problem 2 having a violation range of 17% to 40%; and problem 3 having a violation range of 17% to 45%. There was a significantly greater increase in consistency for discussion groups versus non-discussion groups, but both groups' answers were more consistent on the second presentation.

<u>Slovic and Tversky (1975) study</u>. Slovic and Tversky use the standard Allais problem. Of their 29 college student subjects, 17 chose the Allais response, and 12 chose consistently with Axiom II. On a reconsideration, after reading arguments in favour of each position, 19 subjects chose the Allais response and 10 chose consistently with the axiom. Over the two presentations, 16 made the Allais-type choices, while nine made the axiom-based choices. In a second experiment with 49 student subjects, the subjects first read and rated arguments for and against the axioms, then they made their own choices. With this format, consistency increased. In their actual choices, only 17 subjects made Allais-type choices, but 30 subjects made axiom-based choices. In rating their arguments for the axioms, 25 subjects rated the Allais argument higher while 21 subjects rated the axiom argument higher.

Across these studies, we see that Axiom II is violated at a significant rate. The rate of violation ranges around 27% to 42% except for the high level in Slovic and Tversky's first experiment. However, there seems to be a considerable variation across the studies, and even within a single study.

<u>MacCrimmon and Larsson (1975) study</u>. MacCrimmon and Larsson made a study where the reward set was money, which was varied to determine if a bound exists for which the Allais paradox is not violated. In their study they considered the following variation of the paradox:

	B_1	B_2	B_3	B_4
a	S	S	S	0
b	5S	0	S	0
c	S	S	0	0
d	5S	0	0	0

In this study s took on the values \$1,000,000, \$100,000, \$10,000, and \$1,000 and $\mu(B_1) + \mu(B_2) + \mu(B_3)$ took on the values 1.00, 0.99, 0.50, and 0.11. Eleven different combinations of these parameters were represented plus two check points as a measure of the random component of the choices. For simplicity we shall denote $\mu(B_1) + \mu(B_2) + \mu(B_3) = p_1$ for the choice between a and b, and $p_2 = p_1 - \mu(B_3)$ for the choice between c and d.

From these one can form 12 different sets of two pairs of binary lotteries.

Figures 4.1 provide a summary of the results for the 19 subjects. The tables list the number of subjects making the a and c choice in the higher p set and the b and d choice in the lower p set. For example, in Figure 4.1(i) there were six subjects choosing a and b in the set s = \$1,000,000, $p_1 = 1.00$, and c and d in the set s = \$1,000,000, $p_2 = 0.11$ -- this is the standard Allais problem. Hence the rate of violation is 33% which is about the same as the preceeding studies. Not surprisingly, the highest rate of violation occurs for extreme probability-payoff values. There is not a higher rate of s = \$1,000,000

p_1 \ p_2	1.00	0.99	0.50	0.11
1.00	1	4	4	6
0.99			5	3
0.50				6
0.11				0

Figure 4.1(i)

s = \$100,000

p_1 \ p_2	0.99	0.11
1.00	2	4
0.99		2

Figure 4.1(ii)

s = \$10,000

p_1 \ p_2	0.99	0.11
1.00	4	4
0.99		6

Figure 4.1(iii)

Figure 4.1. Number of subjects inconsistent with utility axioms for various levels of monetary rewards and probabilities.

violation than the standard Allais problem, but two other combinations, $s = \$1,000,000$, $p_1 = 0.50$, $p_2 = 0.11$, and $s = \$10,000$, $p_1 = 0.99$, $p_2 = 0.11$ also have six violations. It should be noted however that these violations are for significantly changed parameter values from the standard problem. They found a significant violation at the \$10,000 payoff level and (separately) at the 0.50 probability level.

4.4 Relations to other axiom systems

This axiom is by far the most important axiom in the sense that empirical evidence shows that it is the one most often violated, and also that the majority of the paradoxes are based on it. Therefore we shall consider the relationship to other axiom systems in some detail.

Considered from alternative approaches, most of the axioms do not show that $f(\cdot)$ may be expressed as a summation of the evaluation function on a subset of Ω. However, they clearly indicate the independence evaluation on a different subset of Ω.

Let us first consider Arrow's approach.

Arrow's Axiom A2 states that if B is an arbitrary event belonging to β, then if two actions have the same rewards on B they must be indifferent given β. This is close to Axiom IIb, the difference being that he does not assume the existence of the function h; however, if h does exist, it must have the properties of Axiom IIb.

Similarly, A4 is related to the first part of Axiom II, although it is not as strong as Axiom II. In this case

Axiom A4 states that if we consider any sequence $\{B_i\}$ of sets in β and if an action a is preferred to an action b on each of these sets (that is if we assume h exists, then $h(B_i,a) \geqslant h(B_i,b)$ for all i) then $f(a) \geqslant f(b)$. This does not imply that h is additive, of course. However, it does imply that the preference on one event does not influence the preference on another event in β.

<u>vonNeumann-Morgenstern's approach</u>. In Appendix I, Axiom NM2 is given as

$$r_1 \overset{R}{>} r_2 \text{ implies } r_1 \overset{R}{>} F(\alpha, r_1, r_2) \text{ and}$$

$$F(\alpha, r_1, r_2) \overset{R}{>} r_2 \text{ for all } \alpha \in (0,1).$$

Intuitively $F(\alpha,r_1,r_2)$ can be considered as receiving r_1 with the probability α and r_2 with the probability $(1-\alpha)$. Let us now consider how this axiom is related to Axiom II.

Consider two reward functions defined by $X(\cdot,a) \equiv r_1$ and $X(\cdot,b) \equiv r_2$ for all $\omega \in \Omega$. Then for any $B \in \beta$, Axiom II states that if action a is preferred to action b then

$$U(r_1) = f(a) = h(B,a) + h(\overline{B},a) \geqslant$$

$$\geqslant U(r_2) = f(b) = h(B,b) + h(\overline{B},b).$$

This clearly implies that at least one of the inequalities $h(B,a) \geqslant h(B,b)$ or $h(\overline{B},a) \geqslant h(\overline{B},b)$ must hold. If we make the assumption that both inequalities hold (Axiom III implies that both must hold), then this implies Axiom NM2. To see this, we shall define a reward function c by

$$X(\omega,c) = \begin{cases} r_1 & \omega \varepsilon B \\ r_2 & \omega \varepsilon \overline{B}. \end{cases}$$

Then if c ε A

$$f(c) = h(B,c) + h(\overline{B},c).$$

By Axiom IIb $h(B,c) = h(B,a)$ and $h(\overline{B},c) = h(\overline{B},b)$ which implies

$$h(B,a) + h(\overline{B},a) \geqslant h(B,c) + h(\overline{B},c) \geqslant h(B,b) + h(\overline{B},b).$$

Savage approach. In this approach there are several related axioms. We shall consider two of these.

Axiom S2: If $X_B(\cdot,a) = X_B(\cdot,b)$, $X_B(\cdot,c) = X_B(\cdot,d)$,

$$X_{\overline{B}}(\cdot,a) = X_{\overline{B}}(\cdot,c),\ X_{\overline{B}}(\cdot,b) = X_{\overline{B}}(\cdot,d),$$

$$\text{and } X(\cdot,a) \overset{\Gamma}{\geqslant} X(\cdot,b),$$

$$\text{then } X(\cdot,c) \overset{\Gamma}{\geqslant} X(\cdot,d).$$

Again if we assume that a real-valued function h on Θ x A exists such that h(B,a) indicates an evaluation of alternative a on the event B, then Savage's Axiom S2 implies that if $X_B(\cdot,c) = X_B(\cdot,d)$ then $h(B,c) = h(B,d)$. This follows directly from Axiom S2. However, it is not necessary that $f(a) = h(B,a) + h(\bar{B},a)$ holds for Axiom S2 to hold. For example $f(a) = h(B,a) \times h(\bar{B},a)$ would also satisfy Axiom S2.

Axiom S3 is also related to Axiom II, where S3 is given by:

<u>Axiom S3</u>: Let $X(\cdot,a) \equiv r_1$ and $X(\cdot,b) \equiv r_2$.
If

$$X_B(\cdot,c) = X_B(\cdot,a),\ X_B(\cdot,d) = X_B(\cdot,b) \text{ and}$$

$$X_{\bar{B}}(\cdot,c) = X_{\bar{B}}(\cdot,d),$$

$$\text{then } X(\cdot,a) \overset{\Gamma}{\leqslant} X(\cdot,b) \text{ if and only if}$$

$$X(\cdot,c) \overset{\Gamma}{\leqslant} X(\cdot,d) \text{ for all } B \subset \Omega$$

such that B is not null.

Axiom S3 is related to Axiom II in the same way as NM2. To see this we shall relate S3 to NM2. Let $X_{\bar{B}}(\cdot,c) = r_1$. Then S3 may be stated as $r_1 \overset{R}{<} r_2$ if and only if $r_1 \overset{\Pi}{<}$ {receiving r_1 if $\bar{B}$ occurs or receiving r_2 if B occurs}. Similarly if $X_{\bar{B}}(\cdot,d) \equiv r_2$ then $r_1 \overset{R}{<} r_2$ if and only if {receiving r_1 if $\bar{B}$ occurs or receiving r_2 if B occurs} $\overset{\Pi}{<} r_2$.

Therefore the relation we have stated between NM2 and Axiom II also holds for S3 and Axiom II.

<u>Marschak approach</u>. In Marschak's approach, Axiom II is the axiom most closely related to Axiom M2. To show this we

can not start with an arbitrary probability space. A special case will be constructed but since this does not relate to the axiom on preference, we feel free to do so.

Let (Ω,Θ,μ) be an arbitrary probability space where Θ is large enough to induce all probability measures in Marschak's approach. Let the probability measures $P(\cdot,a)$, $P(\cdot,b)$ and $P(\cdot,c)$ be induced by $X(\cdot,a)$, $X(\cdot,b)$ and $X(\cdot,c)$ respectively. Let $([0,1],\beta,\lambda)$ be the probability space with β the Borell sets and λ the Lebesque measure, and let $\{[0,1] \times \Omega,\ \beta \times \Theta,\ \mu \times \lambda\}$ denote the product space. We now define two random variables as follows:

$$X((\omega_1,\omega_2),d) = \begin{cases} X(\omega_1,a) & \omega_2 \varepsilon [0,\alpha) \\ X(\omega_1,b) & \omega_2 \varepsilon [\alpha,1] \end{cases}$$

$$X((\omega_1,\omega_2),e) = \begin{cases} X(\omega_1,c) & \omega_2 \varepsilon [0,\alpha) \\ X(\omega_1,b) & \omega_2 \varepsilon [\alpha,1] \end{cases}$$

This implies $f(d) - f(e) = f(a) - f(c)$ by Axiom II since $X_{\Omega x[\alpha,1]}(\cdot,d) = X_{\Omega x[\alpha,1]}(\cdot,e)$ and hence $d \overset{A}{>} e$ if and only if $a \overset{A}{>} c$. Rewriting this into probability measures we have

$$P(\cdot,d) = \alpha P(\cdot,a) + (1-\alpha)P(\cdot,b)$$

$$P(\cdot,e) = \alpha P(\cdot,c) + (1-\alpha)P(\cdot,b)$$

$$\alpha P(\cdot,a) + (1-\alpha)P(\cdot,b) > \alpha P(\cdot,c) + (1-\alpha)P(\cdot,b)$$

$$\text{if and only if } P(\cdot,a) \overset{\Pi}{>} P(\cdot,e).$$

Hence in this sense the axioms are equivalent.

Hagen (1965) criticized this construction in the case of the Allais paradox. In that case our construction would be as follows: Assume actions e,f and g are defined as

$$X(\omega,e) = \begin{cases} \$0 & \omega \in B \\ 5 \times 10^6 & \omega \in \overline{B} \end{cases} \qquad X(\cdot,f) \equiv \$1 \times 10^6$$

$$X(\cdot,g) \equiv \$0$$

and $\mu(B) = 10/11$. Let $\alpha = 11/100$, and define action a, b, c, and d by

$$X((\omega_1,\omega_2),a) = \begin{cases} X(\omega_1,f) & \omega_2 \in [0,11/100) \\ X(\omega_1,f) & \omega_2 \in (11/100,1] \end{cases}$$

$$X((\omega_1,\omega_2),b) = \begin{cases} X(\omega_1,e) & \omega_2 \in [0,11/100] \\ X(\omega_1,f) & \omega_2 \in (11/100,1] \end{cases}$$

$$X((\omega_1,\omega_2),c) = \begin{cases} X(\omega_1,f) & \omega_2 \in [0,11/100] \\ X(\omega_1,g) & \omega_2 \in (11/100,1] \end{cases}$$

$$X((\omega_1,\omega_2),d) = \begin{cases} X(\omega_1,e) & \omega_2 \in [0,11/100] \\ X(\omega_1,g) & \omega_2 \in (11/100,1] \end{cases}$$

Hence we note that actions a, b, c, and d induce the same probability measure on the reward set as in the Allais paradox and clearly

$$X_{\Omega x[0,11/100]}(\cdot,a) \equiv X_{\Omega x[0,11/100]}(\cdot,c)$$

and therefore

$$h(\Omega x[0,11/100],a) = h(\Omega x[0,11/100],c).$$

Similarly

$$h(\Omega x[0,11/100],b) = h(\Omega x[0,11/100],d)$$

$$h(\Omega x(11/100,1],a) = h(\Omega x(11/100,1],b), \text{ and}$$

$$h(\Omega x(11/100,1],c) = h(\Omega x(11/100,1],d).$$

Hence by additivity

$$f(a) - f(b) = f(c) - f(d).$$

If we rewrite this result in terms of probability distributions we have

$$P(\cdot,a) = 11/100\ P(\cdot,f) + 89/100\ P(\cdot,f)$$

$$P(\cdot,b) = 11/100\ P(\cdot,e) + 89/100\ P(\cdot,f)$$

$$P(\cdot,c) = 11/100\ P(\cdot,f) + 89/100\ P(\cdot,g)$$

$$P(\cdot,d) = 11/100\ P(\cdot,e) + 89/100\ P(\cdot,g).$$

Hence $f(a) > f(b)$ if and only if $f(c) > f(d)$ implies

$$11/100\ P(\cdot,f) + 89/100\ P(\cdot,f) > 11/100\ P(\cdot,e) + 89/100\ P(\cdot,f)$$

if and only if

$$11/100\ P(\cdot,f) + 89/100\ P(\cdot,g) > 11/100\ P(\cdot,e) + 89/100\ P(\cdot,g)$$

which is a special case of the strong independence axiom.

Hagen's criticism of this construction -- as we could likewise have defined $X(\cdot,b)$ -- was:

$$X((\omega_1,\omega_2),b) = \begin{cases} X(\omega_1,f) & \omega_2 \in [0,89/100) \\ X(\omega_1,e) & \omega_2 \in [89/100,1) \end{cases}$$

and in this case the induced probability distribution would have been the same; however, Axiom II can not be used to determine a preference between a and b.

Hence we are not stating an exact equivalence between Axiom II and the strong independence axiom. However, we do state that there always exists a probability space and a set of functions defined such that if the strong independence axiom is not found to be acceptable, then neither is Axiom II acceptable. On the other hand, if Axiom II is not acceptable for some functions, then neither can the strong independence axiom be acceptable for the induced probability measure of those functions.

Luce and Krantz approach. In their approach they assume a preference exists on $X_B(\cdot,a)$ for all $B \in \beta$, $a \in A$. Clearly then Axiom IIb must hold (if h exists). Similar to the Savage approach, Axiom LK4 specifies that if $X_D(\cdot,a) \overset{\Phi}{>} X_D(\cdot,b)$ that is, action a is preferred to action b on the set D, then if $D \cap B \neq \emptyset$ and $X_B(\cdot,a) = X_B(\cdot,b)$ then

$X_{D \cup B}(\cdot,a) \overset{\Phi}{>} X_{D \cup B}(\cdot,b)$. This also indicates an independence of evaluating an alternative on different subsets of β. In terms of the function h, we have that $h(D,a) \geqslant h(D,b)$ and $h(B,a) = h(B,b)$ implies that

$$h(D \cup B,a) \geqslant h(D \cup B,b) \quad \text{if } B \cap D = \emptyset$$

As before, however, it does not imply that h is additive.

4.5 Alternatives to Axiom II

There are two basic alternatives to this axiom, either restricting the set of events for which the function h is additive, or increasing this set. The alternatives are then based on the cardinality of set β.

In view of Axioms III and IV, it is implied that there must exist a measure W which is extended from μ on Θ. Hence this would restrict β from a mathematical point of view since the extension may not exist. Hence if we assume for example that β is the set of all subsets of Ω the assumption that a measure exists on β must be relaxed in the same way.

In Savage's (1954) approach he assumed that β is equivalent to all subsets and also that the extended measure is only finitely additive rather than σ-additive.

The other alternative would be to reduce the set for which h is additive, that is there exists a strict subset of β for which h is additive. This would imply that the extension may not be a measure on β. This approach will be discussed in Part II of the thesis.

If we assume for the moment that $\Theta = \{\Omega, \phi\}$, then there exist several alternatives to Axiom II. We shall consider a few of these. All of them assume a function on R called a utility function and then f is specified by the following rules:

The maximax criterion (Hurwitz, 1951). The evaluation function is specified by taking the maximum utility for each action.

The maximin criterion (Wald, 1950). The evaluation function is specified by taking the minimum utility for each consequence.

The Hurwitz α-criterion (Hurwitz, 1951). The evaluation function is specified by a linear combination of the maximum and minimum utility for each action.

There also exist additional alternatives which will be discussed in the introduction to Part II of the thesis.

5.0 Separability axiom

In this section we shall discuss Axiom III. The axiom is formally stated in section 5.1 and its implications are given in section 5.2. Empirical studies are described in section 5.3, its relation to other axiom systems is shown in section 5.4, and finally, the alternatives to it are discussed in section 5.5.

5.1 Statement of Axiom III

In the introduction we stated this axiom as the existence of functions W and U such that $h(r,p) = U(r)W(p)$. The major assertion underlying Axiom III, therefore, is that we can separate the utility of a reward from the probability of receiving the reward. This separability concept was called ethical neutrality by Ramsey (1926), although it was stated slightly differently.

We noted in section 4.2 that if all constant functions from Ω to R belong to A, a utility function may be defined on R. However, if Assumption 2 does not hold, the existence of the function U on R must be assumed. The first part of Axiom III will state the existence of such a function, in addition to a real-valued function W on β. Axiom III then relates the function $h(B,a)$ to the functions U and W.

Before the axiom can be stated, however, we shall need some additional notations and definitions. Informally, a simple function from Ω to the real line E (see Appendix II for definition) is a function which only assigns finitely many values. For example: if $B_1,\ldots,B_n$ are disjoint sets belonging to β whose union is Ω, then $Y(\cdot)$ defined by

$$Y(\omega) = \begin{cases} \alpha_1 & \omega \varepsilon B_1 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \alpha_n & \omega \varepsilon B_n \end{cases}$$

is a simple function.

Let Z be the index set of all simple functions from Ω to E. Then, consistent with our previous notation, we shall let $Y(\cdot,z)$ denote the simple function corresponding to $z \varepsilon Z$. For each function $Y(\cdot,z)$ and for any $B = \bigcup_{i=1}^{k} B_i$ we shall define a number $h(B,z)$ by

$$h(B,z) = \sum_{i=1}^{k} \alpha_i W(B_i),$$

where W is a set function defined in Axiom III.

This notation may, of course, create some confusion since in Axiom II another function $h(B,a)$ was assumed to exist. However, as we shall show, if $UX(\cdot,a)$ is a simple function where $a \varepsilon A$ the two functions are identical.

We also note that if Assumption 2 holds, a complete ordering may be induced on R in the following way:

$r_1 \overset{R}{\geqslant} r_2$ if and only if $U(r_1) \geqslant U(r_2)$ and

$r_1 \overset{R}{>} r_2$ if and only if $U(r_1) > U(r_2)$.

Again, if Assumption 2 is not assumed, an ordering on R can still be specified if U is assumed to exist. Therefore, after we assume the existence of U (Axiom IIIa), whenever we refer to an ordering on R, we mean the induced ordering on R by U.

Axiom III. Separability Axiom

a) There exists a non-negative real-valued σ-additive function W on β and a real-valued measurable function U on R such that for any a ε A, B ε β and r ε R, if $X_B(\cdot,a) = r$, then

$$h(B,a) = W(B)U(r).$$

b) Let r_o be any fixed reward in R, and let $X(\cdot,b)$ be any reward function such that $X_B(\omega,b) \overset{R}{\geqslant} r_o$ for all $\omega \in B$, for B ε β. If Z_o is the set of all simple functions z such that $Y_B(\cdot,z) \leqslant UX_B(\cdot,b)$, then h(B,b) satisfies

$$h(B,b) = \sup_{z \varepsilon Z_o} h(B,z)$$

c) Similarly for any reward function $X(\cdot,c)$ such that $r_o \overset{R}{\geqslant} X_B(\omega,c)$ for all $\omega \in B$ if Z_1 is the set of all simple functions z such that $Y_B(\cdot,z) \geqslant UX_B(\cdot,c)$, then h(B,c) satisfies

$$h(B,c) = \inf_{z \varepsilon Z_1} h(B,z)$$

The first part of the axiom specifies h(B,a) for reward functions which are constant on a set B in β. Since Axiom II implies that h is an additive function, Axiom IIIa also specifies h(B,a) for reward functions which are simple functions. Axiom IIIb and IIIc extend the definition of h(B,a) to an arbitrary reward function belonging to A.

5.2 Implications of Axiom III

First we shall show that if there exists a constant reward function such that $UX(\cdot,a) \not\equiv 0$ then it is not necessary to assume that W is σ-additive since this is implied.

To see this, let $X(\cdot,a) \equiv r$, and B_i, $i=1,2,\ldots,$ be any partition of B, such that $B_i \in \beta$ for all i. We have by Axiom II

$$h(B,a) = \sum_{i=1}^{\infty} h(B_i,a)$$

and by Axiom III,

$$W(B)U(r) = \sum_{i=1}^{\infty} W(B_i)U(r).$$

This implies that if $U(r) \neq 0$, then

$$W(B) = \sum_{i=1}^{\infty} W(B_i).$$

Hence if a constant function $UX(\cdot,a) = U(r) \neq 0$ exists such that $a \in A$ then W must be σ-additive.

For simple functions, Axioms II and IIIa specify the evaluation function f. Let the values of $X(\cdot,a)$ be equal to $r_1,r_2,\ldots,r_n$ and define $B_i = \{\omega : X(\omega,a) = r_i, \omega \in \Omega\}$. Then $\{B_i\}$ is a partition of Ω, and $B_i \in \beta$ for all i. Hence

$$f(a) = \sum_i h(B_i,a) \qquad \text{(by Axiom II)}$$

$$= \sum_i W(B_i)U(r_i) \qquad \text{(by Axiom III)}.$$

Hence in this case we have an expected utility theorem if $W(B_i) = \mu(B_i)$; that is, the evaluation function f is specified by the expected utility of the reward function $X(\cdot,a)$.

We shall also consider some questions of consistency between Axiom II and Axiom III as we must show that the two axioms do not contradict each other. A question arises from the additivity and separability axioms as to whether or not f is uniquely defined. For example if $r_n = r_1$ in our previous example, would f have the same value independently of whether we consider the partition $B_1',\ldots,B_{n-1}'$ (by considering $B_1' = B_1 \cup B_n$, and $B_i' = B_i$ for $i=2,\ldots,n-1$), or the partition $B_1,\ldots,B_n$. It is obvious that f would take the same value if and only if W is an additive set function. By considering a countable partition of any of the sets B_i, the same argument would imply that W must be σ-additive.

In the case where $X(\omega,a) \geqslant r_o$ for all $\omega \in \Omega$ but is not necessarily a simple reward function, we must also show that if $\{B_i\}$ is any arbitrary partition of a set B then Axiom III does not contradict the additivity assumption in Axiom II, that is

$$h(B,b) = \sum_i h(B_i,b).$$

We shall show this by first considering the case when the partition only contains two disjoint sets B_1 and B_2.

If $Y_{B_1 \cup B_2}(\cdot,ac)$ (for a definition, see Appendix III) is a simple function less than $UX_{B_1 \cup B_2}(\cdot,b)$ then $Y_{B_1}(\cdot,a)$ is a simple function less than $UX_{B_1}(\cdot,b)$ and $Y_{B_2}(\cdot,c)$ is a simple function less than $UX_{B_2}(\cdot,b)$. Similarly, if $Y_{B_1}(\cdot,a)$ and $Y_{B_2}(\cdot,c)$ are simple functions less than $UX_{B_1}(\cdot,b)$ and $UX_{B_2}(\cdot,b)$ respectively, then $Y_{B_1 \cup B_2}(\cdot,ac)$ is a simple function less than $UX_{B_1 \cup B_2}(\cdot,b)$. Hence the first assertion implies $\sup h(B_1 \cup B_2,ac) \leqslant \sup h(B_1,a) + \sup h(B_2,c)$ and the second implies $\sup h(B_1 \cup B_2,ac) \geqslant \sup h(B_1,a) + \sup h(B_2,c)$,

Therefore

$$h(B_1 \cup B_2,b) = h(B_1,b) + h(B_2,b).$$

By induction the following equality must be satisfied:

$$h(\bigcup_{i=1}^{n} B_i,b) = \sum_{i=1}^{n} h(B_i,b).$$

To show the general case, we must show that this may be extended to a countable number of sets, that is

$$h(\bigcup_{i=1}^{\infty} B_i,b) = \sum_{i=1}^{\infty} h(B_i,b) \quad \text{for any partition } \{B_i\} \text{ of } B$$

holds for an arbitrary action b.

To show this, we note that

$$\left|h(B,b) - \sum_{i=1}^{k} h(B_i,b)\right| \leqslant |h(B,b) - h(B,z)| \quad +$$

$$\left|h(B,z) - \sum_{i=1}^{k} h(B_i,z)\right| + \left|h\left(\bigcup_{i=1}^{k} B_i,z\right) - h\left(\bigcup_{i=1}^{k} B_i,b\right)\right|$$

by triangle inequality.

Let $h(B,z)$ be the value associated with a simple function such that

$$|h(B,b) - h(B,z)| < \varepsilon/3$$

and

$$\left|h\left(\bigcup_{i=1}^{k} B_i,z\right) - h\left(\bigcup_{i=1}^{k} B_i,b\right)\right| < \varepsilon/3.$$

Since $\left|h(B,z) - \sum_{i=1}^{k} h(B_i,z)\right|$ decreases as k increases, then for k large enough

$$\left|h(B,z) - \sum_{i=1}^{k} h(B_i,z)\right| < \varepsilon/3$$

hence

$$\left|h(B,b) - \sum_{i=1}^{k} h(B_i,b)\right| < \varepsilon$$

for k sufficiently large, therefore

$$h(B,b) = \sum_{i=1}^{\infty} h(B_i,b).$$

What we have shown so far is that Axiom II and Axiom III are consistent with each other; that is, that the definition of $h(B,a)$ does not contradict the additivity assumption in Axiom II for $X(\cdot,a) \geqslant r_o$. Similarly, the same result holds if $X(\cdot,a) \leqslant r_o$.

We also have to show that the number $h(B,\mathbf{b})$ [illegible] is independent of the choice of r_o, i.e., if any other reward r is chosen rather than r_o, $h(B,d)$ is uniquely defined. To do so, let r_1 be any other reward in R, say $r_1 \overset{R}{<} r_o$. Then

$$h(B \cap (X(\cdot,d) \geqslant r_1),a) =$$

$$= h(B \cap (X(\cdot,d) \geqslant r_o),a) + h(B \cap (r_1 \leqslant X(\cdot,d) < r_o),a).$$

Similarly

$$h(B \cap (X(\cdot,d) < r_o),a) =$$

$$= h(B \cap (X(\cdot,d) < r_1),a) + h(B \cap (r_1 \leqslant X(\cdot,d) < r_o),a).$$

Hence

$$h(B \cap (X(\cdot,d) \geqslant r_o),a) + h(B \cap (X(\cdot,d) < r_o),a) =$$

$$= h(B \cap (X(\cdot,d) > r_1),a) - h(B \cap (r_1 \leqslant X(\cdot,d) < r_o),a)$$

$$+ h(B \cap (X(\cdot,d) < r_1),a) + h(B \cap (r_1 \leqslant X(\cdot,d) < r_o),a) =$$

$$= h(B \cap (X(\cdot,d) \geqslant r_1),a) + h(B \cap (X(\cdot,d) < r_1),a).$$

This implies then that any reward $r \varepsilon R$ may be chosen in Axiom III, parts b and c.

So far we have only shown that Axiom III does not violate the previous axioms. We shall now consider the implication of Axiom III by considering some simple decision problems, which have been suggested as "paradoxes" in expected utility theory. In doing so, W is considered a probability, as in section 6.1.

<u>MacCrimmon paradox II (MacCrimmon and Larsson, 1975)</u>. "Two friends on their way to a restaurant decide to order the chef's special of the day although neither knows what it is. On the way in, Tom makes Harry the following offer: Harry is to guess if it will be meat or fish - if he is right, Tom will treat them to a bottle of the best white wine. Harry guesses fish. The wine steward overhears them talking about the wine and tells them that it is out of stock but the best red wine is in stock. Tom then changes the prize to a bottle of red wine. Harry changes his guess to meat."

At first glance, this seems to be a contradiction to Axiom III in the following way: Let $\beta = \{(\text{meat}), (\text{fish}), \Omega, \phi\}$ where (meat) = (the event that meat is the chef's special) and similarly for (fish). Harry guessing fish implies $U(\text{white wine})W\{(\text{fish})\} > U(\text{white wine})W\{(\text{meat})\}$. If $U > 0$ then $W\{(\text{fish})\} > W\{(\text{meat})\}$ and therefore

$$U(\text{red wine})W\{(\text{fish})\} > U(\text{red wine})W\{(\text{meat})\}.$$

Hence by revising his guess this would indicate that he thought meat was more likely. That is $W\{(\text{meat})\} > W\{(\text{fish})\}$ contradicts our previous conclusion. Harry's behaviour may be perfectly rational, however, in that we may be defining the reward set incorrectly. Harry does not contradict Axiom III if our definition of the reward set is equal to (white wine with fish), (white wine with meat), (red wine with fish), (red wine with meat).

Since Tom offered white wine either (1) the natural association would be fish so Harry said that, or (2) since Tom offered white wine this implied that he expected fish so Harry guessed accordingly, or he changed to meat because he thought it unlikely the chef would put on a fish special when he was out of white wine.

Another paradox apparently contradicting the separability part of the axiom is called the Newcombe paradox.

Newcombe paradox - Nozick (1969). Consider the following situation: Two closed boxes A and B are on the table in front of you. Box A contains $1,000. Box B contains either nothing or $1,000,000. You do not know which. You have a choice between two actions:

(i) Take what is in both boxes,

(ii) Take only what is in box B.

At some time before this opportunity, a superior being made a prediction about what you will decide. The being is "almost certainly" correct. If the being expects you to take action (i), he will leave box B empty. If he expects you to take action (ii), he will leave $1,000,000 in box B. If he expects you to randomize your choice, for example by flipping a coin, he will leave box B empty. In all cases, box A contains $1,000.

Which action would you choose? Consider the following arguments: (i) Either the money is in box B or it is not. If the money is in box B and I take both boxes, I will have $1,000 more than if I had only taken Box B. Alternatively, if the money is not in box B, and I take both boxes, at least I will get $1,000. Hence, taking alternative (i), i.e., selecting both boxes, is the better strategy. (ii) If the being can guess with "almost certainty" then I would only take box B, since if I were to take both, he would almost surely guess correctly and hence leave box B empty.

This problem, and its associated arguments, was first published by Nozick (1969) and is called the "Newcombe paradox". The Newcombe problem differs from others in several ways. Perhaps most importantly, it presumes to set one of the axioms in opposition to the expected utility criterion, rather than to attack one of the axioms with a counter-axiom. Argument (i) is based on the dominance axioms, that is, if alternative (i) is always better than alternative (ii) independently of what state of nature occurs, choose (i). In symbols we have that if $X(\omega,a) < X(\omega,b)$ for all $\omega \in \Omega$ then $a < b$. The argument for alternative (ii) is based on an expected utility formulation. Presumably we cannot have both.

It is useful to analyse more directly how the dominance and expected utility formulations apparently contradict each other. Let us look first at dominance, as expressed most directly in Arrow's Axiom A4. Dominance is almost universally accepted as a reasonable axiom to use when it applies, and so it would be hard to choose in contradiction to it. Consider the following way of formulating the problem in a payoff matrix:

		$1,000,000 in box B	Nothing in box B
(i)	Take both boxes	$1,001,000	$1,000
(ii)	Take only box B	$1,000,000	$0

If we accept this formulation of the problem, it would be difficult not to take alternative (i) because it dominates alternative (ii).

However, this formulation may be questionable because it fails to take into account the predictive ability of the superior being. Consider, instead, the following payoff matrix formulation:

		Being predicts correctly	Being does not predict correctly
(i)	Take both boxes	$1,000	$1,001,000
(ii)	Take only box B	$1,000,000	$0

Obviously in this formulation dominance does not apply and one would take alternative (ii) if P (being correct) = p, and

$$pU(\$1,000,000+(1-p)U(\$0) > pU(\$1,000)+(1-p)U(\$1,001,000).$$

For any reasonable utility function, and assuming that p is close to 1 as implied in the problem, alternative (ii) would have the higher expected utility. Thus, if it were not for the different formulations, one would have the paradoxical situation in which dominance implies one action while the maximization of of the expected utility implies another action.

The major difference, then, between this challenge to the axioms and those considered earlier, is that the "Newcombe paradox" is based on the way the problem is formulated, rather than in the choices offered in a specific formulation. Expected utility theory requires an independence between the events and the actions. In the "dominance formulation" of the problem, the probability of either event occurring is not independent of our choice of actions and is therefore inappropriate. While this difficulty does not hold for the second formulation above, the second formulation does not take into account the amount in the boxes and hence may seem incomplete. In order to get both uncertain elements into the problem, we need to form the compound events:

		Being predicts correctly and put:		Being predicts incorrectly and put:	
		$1,000,000 in box B	$0 in box B	$1,000,000 in box B	$0 in box B
(i)	Take both boxes	-	$1,000	$1,001,000	-
(ii)	Take only box B	$1,000,000	-	-	$0

The crossed out cells represent impossible combinations and an examination of the whole table shows that dominance cannot be applied. Hence one can choose only box B and act in accordance with expected utility without violating dominance. If there are very large, non-monetary satisfactions of exhibiting the "free-will" of taking both boxes or of beating the being out of \$1,001,000 and showing him up in the process, as asserted by Asimov (Gardner, 1974, p.123), then you might choose action (i). You would, however, be choosing it on an expected utility basis rather than on the basis of dominance.

Allais paradox II (Allais, 1953). There exists a simple generalization of the Allais paradox I. Consider, for example, two functions defined as $X_B(\cdot,a) = r$, and $X_D(\cdot,b) = s$ where $h(B,a) = h(D,b)$. This implies

$$U(s)W(D) = U(r)W(B)$$

or equivalently

$$U(s) = \frac{W(B)}{W(D)} U(r), \quad \text{if } W(D) \neq 0.$$

Then, for any other functions $X_E(\cdot,c) = r$ and $X_F(\cdot,d) = s$ such that $\frac{W(D)}{W(F)} = \frac{W(B)}{W(D)}$ it is implied that $h(E,c) = h(F,d)$. This gives rise to the paradox of common ratio. With Axiom IV we shall see that $W(B) = \mu(B)$ for all $B \in \Theta$, and we will use this assumption here to illustrate the common ratio paradox. If we assume that the reward set is the real line and that $U(0) = 0$, then the paradox of common ratio of probabilities implies that:

A preference of a to b implies a preference of c to d where a, b, c, and d are defined as follows:

a: \$1 million with a probability of 1.0
\$0 otherwise
b: \$5 million with a probability of 0.8
\$0 otherwise

c: \$1 million with a probability of 0.05
\$0 otherwise

d: \$5 million with a probability of 0.04
\$0 otherwise

Since 0.8/1 = 0.04/0.05, the ratio of probabilities is common.

This therefore implies that if $f(a) > f(b)$ then $f(c) > f(d)$. Empirical studies show that this is not always so, in that people commonly select a and d.

5.3 Empirical studies on Allais' paradox II

In the two paradoxes considered in the preceding section (i.e., MacCrimmon and Newcombe), the issue revolved around a definition of the problem rather than an empirical implication of the axiom. Thus empirical studies of these problems would not give any further support for or against the axiom. Hence we shall only consider empirical studies of the common ratio problem. The paradox is usually written in the following form. Let the reward space be the real line, and let α and β be two real numbers.

A choice is to be made between alternatives a and b:

a: receiving r with a probability of p

b: receiving αr with a probability of βp.

The same preference must then hold for all values of p. Let us therefore choose two values of p, p_1 and p_2 and for simplicity we shall call the first choice in the above problem a_1 or b_1, and the second choice a_2 or b_2, and hence a preference $a_1 \overset{A}{>} b_1$ must imply $a_2 \overset{A}{>} b_2$.

<u>Hagen's (1971) study</u>. Hagen obtained some evidence from Norwegian teachers when he used the problem with the following

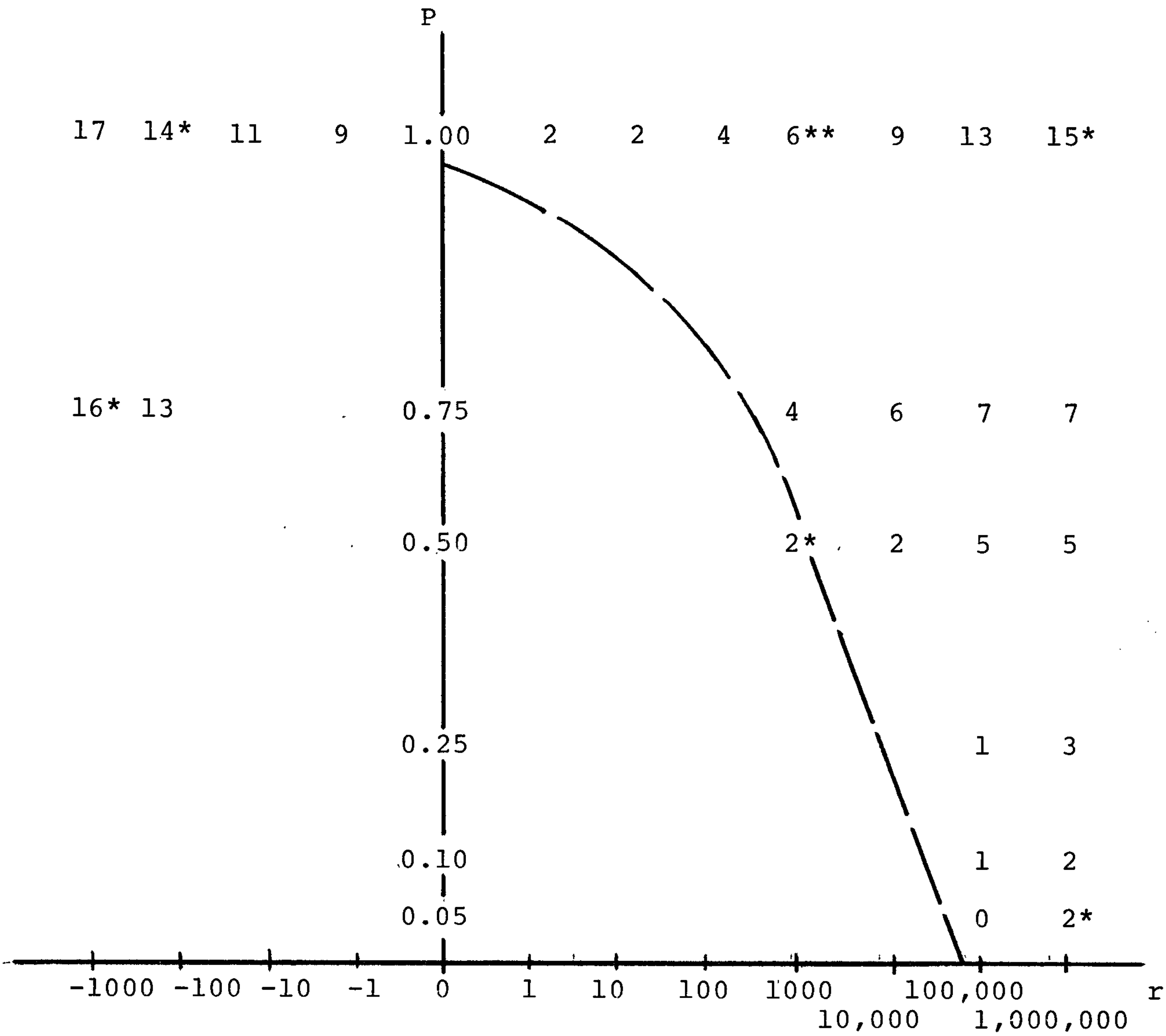
parameter values: $r = 1$ million Norwegian kroner, $\alpha = 5$, $p_1 = .99$, $p_2 = .11$, and $\beta = 10/11$. Because subjects were asked for choices in only one of the sets, Hagen could compare only the aggregate number of a vs. b choices in the two sets; he could not compare each individual subject's choices across both sets. Hagen found that in the first set, 37 subjects out of 52 selected a_1 while in the second set 37 subjects of 52 selected b_2. Hence there was a preference for the a_1 alternative in the first set but the b_2 alternative in the second set. Thus, we can infer that if subjects had been presented with both sets, the majority would probably have violated the expected utility hypothesis. In a second experiment, Hagen used the parameter values: $r = 10{,}000$ Norwegian kroner, $\alpha = 2$, $p_1 = 1.00$, $p_2 = 0.02$, and $\beta = 1/2$. In the first set of choices, 47 of the 52 subjects selected a_1 while in the second set 23 of the 52 subjects selected b_2. Even though the a_2 alternative was the more favoured one in the second set, this pattern of choices again suggested some violation of Axiom III.

MacCrimmon and Larsson (1975). MacCrimmon and Larsson attempted to study the effect of varying some of the parameter values concentrating on the effect of using different payoff and probability levels (i.e., values of r and p). They used two experiments. In the first they used positive payoffs; in the second they used negative payoffs (losses) to determine whether negative payoffs resulted in major differences in behaviour. The payoffs were all hypothetical but the subject was asked to act as if each would actually be realised and to treat each set independently of the others. The parameter values used were the following:

On the positive expected value sets: $\alpha = 5$, $\beta = 4/5$, r took on the values \$1,000,000; \$100,000; \$10,000; \$1,000; \$100; \$10; and \$1, while p took on the values 1.00, 0.75, 0.50, 0.25, 0.10, and 0.05. On the negative expected value sets: $\alpha = 5$, $\beta = 3/4$, r took on the values -\$1,000; -\$100; -\$10; and -\$1, while p took on the values 1.00, 0.80, 0.20, and 0.04.

Twenty-one different positive expected value sets were presented, with four sets repeated to check for consistency. Eight different negative expected value sets were presented with two sets repeated to check for consistency. Not all combinations were presented since a pilot study had ascertained that some combinations (e.g., a low positive payoff and a low probability level) led to almost all subjects choosing the same alternative. (This is interesting in itself, but is not the best use of limited time for an experiment.) The 25 positive payoff sets and 10 negative payoff sets were presented in random order.

The particular combinations given can be seen from the graph of the results in Figure 5.1. The numbers show for each payoff-probability combination how many of the 19 subjects chose the higher probability (a) alternative. So, for example, with the combination of payoffs and probabilities used in section 5.2 above, 15 subjects chose a in set 1 while only two chose a_2 in set 2. Not surprisingly, the highest level of a choices, for positive amounts, occurs when there is a sure



* One subject chose a_1 on one presentation of this set, and b_1 on the other presentation.

** Two subjects chose a_1 on one presentation of this set and b_1 on the other presentation.

Fig. 5.1. The number of subjects, out of 19, selecting the contradictory preferences to expected utility criterion in the problem of Common Ratio of Probabilities

chance of getting a large amount of money. As the graph shows, when the probability levels decrease, or when the money payoff levels decrease, then there is a reduced tendency to pick the a alternative (i.e., the one giving the lower payoff with the higher probability). When the probability levels decrease the rationale is one of viewing the probability difference as insignificant and thus "going for broke" on the larger payoff. When the payoff levels decrease, the rationale is one of "going for broke" since the amount you get for sure does not mean that much to you in terms of lifetime security, etc. The majority of the subjects would only select the a alternative when there was a certainty of getting a very large payoff (i.e., either $1,000,000 or $100,000). Even though each of our subjects made 25 (positive payoff) choices, they seemed to be quite alert to the changes in payoff and probability and hence chose differentially. It is clear, then, that the particular parameter values play a major role in whether one violates the utility independence conditions. Since almost all subjects can be expected to prefer the b alternative for payoff-probability combinations to the left and below the dashed line, then there would be no violation of the expected utility criterion, if any of these combinations were compared to each other. Since in real choices the subjects would rarely have alternatives with payoffs such that they would be to the right of the line, we may question whether the possible violations in very unlikely cases have much relevance for utility theory.

Some countervailing evidence, though, is found by examining the negative payoffs. In this case there is more

α = \$1,000,000

p	1.00	.75	.50	.25	.10	.05
1.00	1*	6	9	11	11	12
.75			7	5	5	6
.50				2	2	3
.25					0	1
.10						1
.05						1*

(i)

α = \$1,000

p	1.00	.75	.50
1.00	2*	6	4
.75			4
.50			1*

(ii)

α = −\$1,000

p	.80	.20	.04
1.00	1	1	3
.80	1*	2	4
.20			4

(iii)

Fig. 5.2 Number of subjects is consistent with Axiom III for different values of α-p in variations of Allais paradox II.

*These subjects were inconsistent on the repeat of the same set, hence they were omitted from the tabulation of the remainder of the table. So the inconsistencies in Table (i) are out of 17 subjects, in Table (ii) out of 16 subjects, and in Table (iii) out of 18 subjects.

ambivalence in switching from a to b as the probability level or size of the loss decreases since the loss levels were chosen to be ones that would be realistic for the subjects. The choices that subjects made across two sets are shown for various combinations in Figure 5.2. The value 12 in the upper-right corner of Figure 5.2(i) tells us, for example, that when the choices of a subject for the probability level 1.00 are compared to his choices at the 0.05 level, 12 of the subjects chose a in one of the sets and b in the other, hence violating Axiom III. Note that those direct comparisons confirm what we observed in Figure 5.1, that large payoff and probability levels lead to a higher propensity for violating Axiom III. Note though that the negative payoff results indicate a relatively low level of violation.

5.4 Relation to other axiom systems

In Axiom III we state two properties, first, the separability between the rewards, and second, the method of evaluating $h(B,a)$. In comparing these concepts to other previously mentioned systems we encounter some difficulties since nearly all these axioms are needed to evaluate $h(B,a)$. However the first property of separability is easier in some cases to compare and we shall do so here.

In Axiom IIb we assumed that if $X_B(\cdot,a) = X_B(\cdot,b)$ then $h(B,a) = h(B,b)$. That is, if the two functions are identical on an event, then the evaluation for the actions on that event must be the same. The separability assumptions extend this idea to the case where if $X(\cdot,a)$ is a constant function on the event

$B \in \beta$, and $X(\cdot,b)$ has the same constant value on $C \in \beta$ then

$$h(B,a) = h(C,b) \quad \text{if and only if}$$

$$W(B) = W(C).$$

Similarly if $X_B(\cdot,a) = r_1$ and $X_B(\cdot,b) = r_2$ are constant reward functions on B

$$h(B,a) = h(B,b) \quad \text{if and only if}$$

$$U(r_1) = U(r_2).$$

In each of the approaches of vonNeumann & Morgenstern, Marschak, and Arrow it is assumed that the probabilities are given and we ignore the events given to these probabilities, i.e., this implies that $h(B,a)$ is of the form $h(\mu(B),a)$. This is clearly a separation between the event and its reward, since we are only evaluating $\mu(B)$, and would be indifferent to any other event D with the same reward for which $\mu(D) = \mu(B)$. However this does not necessarily imply that the probability and the reward may be separated. In their approaches they imply that $h(B,a) = h(C,b)$ if and only if $W(B) = W(C)$. To see this consider the case where $X_B(\cdot,a) \equiv r$, $B \in \beta$, and $W(B) = \alpha$, then Axiom III implies that

$$h(B,a) = \alpha U(r).$$

If γ is any number between (0,1), then

$$\gamma h(B,a) = \alpha\gamma U(r)$$

or equivalently $\gamma h(B,a)$ must equal the evaluation of any action c for which $X_D(\cdot,c) = r$ where $W(D) = \alpha\gamma$. This in turn implies that if $X_B(\cdot,a) \equiv r$, $X_E(\cdot,c) \equiv r$, $B,E \in \beta$, $B \cap E = \emptyset$, $W(B) = \alpha$, $W(E) = \gamma$. Then $(1-\delta)h(B,a) + \delta h(E,c)$ must be equivalent to an action d defined by $X_D(\cdot,d) \equiv r$ and $W(D) = (1-\delta)\alpha + \delta\gamma$. In more general terms this implies that if $P(\cdot,a)$, $P(\cdot,b)$ and $P(\cdot,c)$ are three probability measures belonging to Π such that the mathematical identity

$$\delta P(\cdot,a) + (1-\delta)P(\cdot,b) = P(\cdot,c)$$

holds we must also have the preference

$$\delta P(\cdot,a) + (1-\delta)P(\cdot,b) \overset{\Pi}{=} P(\cdot,c).$$

This preference is expressed in both Marschak's and vonNeumann & Morgenstern's approaches.

There have been some criticisms against this for the following reason. Suppose we are offered a prize if a red ball is drawn from either urn I or urn II, of our choice. Urn I contains 50 black and 50 red balls. Urn II has been drawn at random from a collection of 101 urns, one of which had 0 red and 100 black balls, another had 1 red and 99 black balls and so on up to one having 100 red and 0 black balls. If we assume that the probability of a red ball being drawn from any urn is equal to the number of red balls divided by the total number of balls in the urn, we must be indifferent as to the choice of urns since

$$1/101 \cdot 0/100 + 1/101 \cdot 1/100 + 1/101 \cdot 2/100 + \ldots + 1/101 \cdot 100/100 = 1/2.$$

However once the second urn has been chosen, there is clearly a fixed number of red balls in that urn, and the preference may not be exactly the same since the chances in urn II are not exactly known. For example, if urn II were chosen at random from two urns, one with 100 red balls and one with 0 red balls, ought this to be equivalent to an urn with 50 red and 50 black balls when we know that urn II in front of us can only have either 0 or 100 red balls? The argument that urn II contains 50 red balls on the average can clearly not be used since this implies that we have the choice repeatedly rather than once.

In Arrow's approach he assumes that if two probability distributions are equal they must have the same preference. Hence this is equivalent to saying that if

$$\alpha P(\cdot,a) + (1-\alpha)P(\cdot,b) = P(\cdot,c)$$

the preference must also be the same which we have already discussed.

In Savage's approach the separability implications are made by changing the reward on the set. Recall Savage's Axiom

S4:

<u>Axiom S4</u>: If $B, C \subseteq \Omega$ and $r_i \in R$ for $i=1,2,3,4, r_1 > r_2$, $r_3 > r_4$ then for actions $a,b,c,d \in A$, defined by the reward functions

$$X_B(\cdot,a) = r_1 \quad X_{\overline{B}}(\cdot,a) = r_2 \quad X_C(\cdot,b) = r_1 \quad X_{\overline{C}}(\cdot,b) = r_2$$

$$X_B(\cdot,c) = r_3 \quad X_{\overline{B}}(\cdot,c) = r_4 \quad X_C(\cdot,d) = r_3 \quad X_{\overline{C}}(\cdot,d) = r_4$$

and if $X(\cdot,a) \overset{\Gamma}{\leqslant} X(\cdot,b)$, then $X(\cdot,c) \overset{\Gamma}{\leqslant} X(\cdot,d)$.

Assume that $r_2 = r_4$, then $h(\overline{B},a) = h(\overline{B},c)$ and $h(\overline{C},b) = h(\overline{C},d)$. If

$$h(B,a) + h(\overline{B},a) < h(C,b) + h(\overline{C},b)$$

for a given $U(r_1) > U(r_2)$ then the same equality must be true for $r \in R$ such that $U(r_3) > U(r_2)$. This implies that

$$h(B,c) < h(C,d) + \text{const}$$

if $U(r_3) > U(r_2)$ where $X_B(\cdot,c) = X_C(\cdot,d) = r_3$. That is, the preference can be determined by comparing if the reward is above a fixed reward.

5.5 Alternative to Axiom III

Axiom II assumes $W(B)$ is a real-valued function on β which implies that if two reward functions are defined by

$$X_B(\cdot,a) = r_1 \text{ and } X_B(\cdot,b) = r_2$$

then

$$h(B,a) > h(B,b) \text{ if and only if } U(r_1) > U(r_2).$$

For example, say that we are invited to dinner where we know that either chicken, beef or fish is to be served, and we decide to bring a bottle of wine, either red, white, or rosé. We also assume the following utility of the reward:

	Chicken	Beef	Fish
a	1	-1	1
b	0	1	-1
c	0.5	0	-1

where action a is to bring a bottle of white wine, action b is to bring a bottle of red wine, and action c is to bring a bottle of rosé wine.

Therefore if B is the event beef is served we have

$$h(B,a) = -1W(B)$$
$$h(B,b) = 1W(B)$$
$$h(B,c) = 0W(B)$$

and hence the preference may be determined by comparing -1, 1, and 0. The assumption here is that W(B) is independent of our action. That is if we construct a matrix indicating the values of W(B) for each action we would have

	Actions		
	a	b	c
Chicken	W(C)	W(C)	W(C)
Beef	W(B)	W(B)	W(B)
Fish	W(F)	W(F)	W(F)

The value f(a) can then be found by multiplying the row a of the reward matrix by column a in event matrix, i.e.,

$$f(a) = h(C,a) + h(B,a) + h(F,a)$$

$$= 1W(C) + (-1)W(B) + 1W(F),$$

and similarly for f(b) and f(c).

This example is taken from R. C. Jeffrey's book "The logic of decision" (1965) where he develops a theory by arguing that the evaluation of W(B) ought to be a function from $\beta \times A$ rather than only β. Hence the event matrix can take the form

	a	b	c
Chicken	k_1	k_2	k_3
Beef	l_1	l_2	l_3
Fish	m_1	m_2	m_3

where k_i, l_i and m_i $i=1,2,3$ are non-negative real-numbers such that

$$k_i+l_i+m_i = 1 \quad i=1,2,3.$$

It is easy to see that the Newcomb paradox falls into this type of decision problem. The reward matrix would be as before,

	Being predicts correctly	Being does not predict correctly
a	U($1,000)	U($1,001,000)
b	U($1,000,000)	U($0)

and the event matrix would be

	a	b
Being predicts correctly	P_a	P_b
Being does not predict correctly	$(1-P_a)$	$(1-P_b)$

where P_a = probability that the being predicts correctly given action a is chosen.

Hence

$$f(a) = U(\$1,000)P_a + U(\$1,001,000)(1-P_a)$$
$$f(b) = U(\$1,000,000)P_b + U(\$0)(1-P_b),$$

and therefore the preference would depend on P_a and P_b.

There is some behavioural support for the concept that W(B) also depends on the action. For example, when betting during a game of roulette, some people would argue that they are always unlucky and will therefore lose, while others are always lucky and will win. That is, the probability of winning does not only depend on the ivory ball and the roulette wheel but also on who does the betting.

6.0 Probability axiom

First, we shall state the axiom in section 6.1. In section 6.2 we shall consider the implications of the axiom and in section 6.3 we shall explore some of the empirical evidence regarding the paradoxes related to this axiom. In section 6.4 we shall compare this axiom to other systems, and finally in section 6.5 we deal with some alternatives to this axiom.

6.1 Statement of Axiom IV

In the introduction Axiom IV was stated as $W(B) = \mu(B)$. In section 5, W was specified as a measure on a σ-algebra β containing Θ. Since μ is only defined on Θ, W may be thought of as an extension of the measure μ on Θ to a measure W on β.

The previous axiom is sufficient to specify the evaluation function as

$$f(a) = \int UX(\omega, a)\,dW.$$

Hence the evaluation function is specified as the expected utility. However, for $f(\cdot)$ to have some meaning for the decision maker, W must in some way be connected to the probability of the different states occurring. It is easily seen that if $W(B)$ is defined by $W(B) = a\mu(B)$ $a > 0$, W would still be a σ-additive measure and the expected utility calculated by using W would give the same ordering for all positive values of a. For simplicity rather than necessity, we shall assume that a = 1.

<u>Axiom IV</u>. For any $B \in \Theta$, $W(B) = \mu(B)$.

Axiom III assumes the existence of the set function W on β. However, Axiom III does not give us a method of determining the values of W(B). Axiom IV specifies those values for all $B \in \Theta$, and it also gives the required condition of specifying W for all B in β. How this is done, we shall discuss in section 6.2.

Very few empirical studies have been made to verify the fact that people act as though $\mu(B) = W(B)$. Of course, the axiom can not be tested directly, and it is rather the people's ability to estimate W(B) for different values of $\mu(B)$ which is tested, or to see if W is a measure. We shall discuss how this is done in more detail in section 6.3. In general, those studies which have been made indicate that W(B) is overestimated if $\mu(B)$ is "small" and underestimated if $\mu(B)$ is "large".

6.2 Implications of Axiom IV

There have been some suggestions that Axiom IV should not always be satisfied. One of these critics, Menger (1950), has suggested that sets with "small probabilities" are to be regarded as impossible. This suggestion creates as many difficulties as it solves. The main difficulty which arises is the meaning of "small probabilities" although some definitions exist. In statistics a decision criterion is often used which ignores probabilities up to 0.05. Similarly in chance-constrained programming we are often willing to ignore small probabilities, usually 0.05 or less. Menger's suggestions have some validity, based on empirical studies. Consider for example, the decision problem in section 5.2 where the alternatives were considered as follows:

a: \$1 million with a probability of 1.0
\$0 otherwise

b: \$5 million with a probability of 0.80
\$0 otherwise.

Thus, the difference of winning between the alternatives is a probability of 0.20. However, when we compare alternatives c and d where

c: \$1 million with a probability of 0.05
\$0 otherwise

d: \$5 million with a probability of 0.04
\$0 otherwise

that the difference is only 0.01. One argument which has been suggested for choosing d rather than c is that the difference in the probabilities of winning is "small" enough to ignore especially since the probabilities of winning are very small. In the case of a vs. b the difference in the probabilities of winning is 0.20, too "large" to ignore, and hence we may choose a. This would imply that W is not a linear function of μ, (or equivalently W is not additive) and hence this preference does not support the expected utility criterion.

The most important aspect of Axiom IV is its usefulness in deriving the values for W(B) where $B \in \beta - \Theta$. The method typically used is as follows:

If two actions $a, b \in A$ are defined according to the

$$X(\omega,a) = \begin{cases} r & \omega \in B \\ s & \omega \in \bar{B} \end{cases} \qquad X(\omega,b) = \begin{cases} r & \omega \in D \\ s & \omega \in \bar{D} \end{cases}$$

where $U(r) > U(s)$, then

$$f(a) = U(r)W(B) + U(s)\ (1-W(B))$$

$$f(b) = U(r)W(D) + U(s)\ (1-W(D))$$

or equivalently

$$f(a) - f(b) = U(r)\left[W(B) - W(D)\right] - U(s)\left[W(B) - W(D)\right]$$

$$= \left[U(r) - U(s)\right]\left[W(B) - W(D)\right].$$

Since by assumption $U(r) > U(s)$, it follows that $f(a) > f(b)$ if and only if $W(B) > W(D)$. If it is assumed that for every real number $\gamma \in (0,1)$ there exists a $D \in \Theta$ such that $\mu(D) = \gamma$, then it is easily seen that $W(B)$ can then be estimated as closely as we wish if A contains sufficiently many comparable actions, in the following manner:

Action a is compared to action b:

1) if $a \overset{A}{\geqslant} b$ $\qquad \gamma \geqslant W(B)$

2) if $b \overset{A}{\geqslant} a$ $\qquad W(B) \geqslant \gamma$.

If case 1 holds a smaller γ can be selected, and a new comparison is made. Similarly for case 2, the process is repeated until W(D) has been estimated to the degree of accuracy desired.

Some difficulties arise, however, as we attempt to extend μ to β. We shall discuss some of them below. The first is strictly mathematical and is concerned with the existence of the extension of μ. For example, if β is all subsets of Ω, there may not exist a measure on β which is equivalent to μ on Θ. Hence we must make the additional assumption that this extension can always be made. This implies that we must restrict the set β or equivalently restrict the number of actions in A.

For practical purposes this can usually be done, since β can in general be generated by Θ and a finite number of sets and in this case the extensions always exist (see section 8 in Part II).

A second difficulty arises in the attempt to determine the value of W(B). One reason is due to the assumptions of separability between the probabilities of receiving a reward and the reward. This can be best illustrated by an example in DeGroot's (1970) book.

Consider actions a and b where:

a: Receiving \$100 if you will be exterminated by a nuclear war within the next ten years, or \$0 otherwise

b: Receiving \$100 if you become the president of the United States within the next ten years, or \$0 otherwise.

It is not surprising that most people would prefer b to a and also believe that the event {extermination by a nuclear war} is more likely to occur than the event {becoming the president of the United States within the next ten years}. The problem arises because the rewards are not precisely defined or exactly specified. If rewards are not receiving $100 but rather "receiving $100 and being exterminated" and "receiving $100 and being president" the contradiction will not occur. In this case an obvious relation exists between the rewards and the events yielding them. In some cases, however, it may be possible to determine if such a relation exists.

Another difficulty arises when we attempt to determine the value of W(B). Assume that W(D) has been determined to equal γ. This value may not then satisfy the additivity property of a measure.

Let us illustrate this using the Ellsberg paradox I.

Ellsberg paradox I. Consider the following two urns: Urn I contains 100 balls, either red or black though the number of each colour is not known. Urn II contains 50 red balls and 50 black balls. We are asked to state a preference between a and b and a preference between c and d.

a: Win $1,000 if a red ball is drawn from Urn I

b: Win $1,000 if a red ball is drawn from Urn II

c: Win $1,000 if a black ball is drawn from Urn I

d: Win $1,000 if a black ball is drawn from Urn II.

Let us denote the event (drawing a red ball from Urn I) by $\{R_I\}$ and similarly for $\{R_{II}\}$, $\{B_I\}$ and $\{B_{II}\}$. Hence if someone strictly prefers b to a we can conclude that

$W(R_{II}) > W(R_I)$. Similarly, if our preference is $d \overset{A}{>} c$, we must have $W(B_{II}) > W(B_I)$. If both preferences $b \overset{A}{>} a$ and $d \overset{A}{>} c$ are made, a contradiction occurs, i.e.,

$$1 = \mu(\Omega) = W(R_{II}) + W(R_{II}) > W(R_I) + W(B_I) = \mu(\Omega) = 1.$$

Thus, the additivity of W must be rejected and hence if our preference is $b \overset{A}{>} a$ then we must also have the preference $c \overset{A}{>} d$.

This problem can also be illustrated by our attempt at specifying $W(R_I)$. Assume that the only events that belong to Θ are $\Omega, \Theta, (R_{II})$, and (B_{II}), and also that if there is a proportion p of red balls in Urn II, the probability of drawing a red ball is p. W(R) can now be determined by comparing the actions:

a: Receiving \$1,000 if a red ball is drawn from Urn I,

c: Receiving \$1,000 if a red ball is drawn from an urn with the proportion p of red balls.

When the decision maker becomes indifferent between a and b we let $W(R_I) = p_1$. If we repeat our experiment for black balls we shall find $W(B_I) = p_2$. Unfortunately most empirical studies indicate that $p_1 + p_2 \neq 1$. It is not easy to determine if p_1 or p_2 or both should change.

A fourth difficulty arises if Θ does not contain a sufficient number of events such that $W(B)$ $(B \in \beta - \Theta)$ can be estimated with sufficient accuracy by the method suggested. An alternative approach exists if β contains a sufficient number of sets. Consider again the two actions $a, b \in A$, defined by

$$X(\omega,a) = \begin{cases} r & \omega \in B \\ s & \omega \in \overline{B} \end{cases} \qquad X(\omega,b) = \begin{cases} r & \omega \in D \\ s & \omega \in \overline{D} \end{cases}$$

where $U(r) > U(s)$. An ordering can then be defined on the sets in β in the following way:

If $a \overset{A}{\leqslant} b$ then $B \overset{\beta}{\leqslant} D$.

If the ordering $\overset{\beta}{\leqslant}$ satisfies a given set of assumptions, it can be proved that there exists a real-valued function P on β, which satisfies the axioms of a probability measure. (For proof see, for example, De Finetti, (1964) or Savage (1954)).

We shall summarize the conditions on the ordering $\overset{\beta}{\leqslant}$ for the existence of P below:

The relation ("is not more probable than") on events must satisfy the following axioms:

<u>Axiom of ordering</u>.

1. If $C \in \beta$, $B \in \beta$, then either $C \overset{\beta}{\leqslant} B$ or $B \overset{\beta}{\leqslant} C$;
2. For any set $B \in \beta$, $B \overset{\beta}{\leqslant} B$.
3. If $D \overset{\beta}{\leqslant} B$ and $B \overset{\beta}{\leqslant} C$, then $D \overset{\beta}{\leqslant} C$.
4. $\phi \overset{\beta}{<} \Omega$ and for any event $B, \phi \overset{\beta}{\leqslant} B \overset{\beta}{\leqslant} \Omega$.

We define $\overset{\beta}{<}$ ("is strictly less probable than") in the usual way by:

$$C \overset{\beta}{<} B \text{ implies } C \overset{\beta}{\leqslant} B \text{ and not } B \overset{\beta}{\leqslant} C.$$

<u>Axiom of monotonicity</u>.

1. If $B_1 \cap B_2 = \emptyset$, $C_1 \overset{\beta}{\leqslant} B_1$ and $C_2 \overset{\beta}{\leqslant} B_2$, then

$$C_1 \cup C_2 \overset{\beta}{\leqslant} B_1 \cup B_2$$

2. If $B_1 \cap B_2 = \emptyset$, $C_1 \overset{\beta}{\leqslant} B_1$ and $C_2 \overset{\beta}{<} B_2$, then

$$C_1 \cup C_2 \overset{\beta}{<} B_1 \cup B_2.$$

The relation $\overset{\beta}{\leqslant}$ satisfying the above axioms is called a qualitative probability on the algebra β. Savage's axioms 1-5 imply these axioms and in his assumptions the algebra β contains all subsets of Ω. However, these axioms are not sufficient to guarantee σ-additivity of the corresponding probability measure. For σ-additivity we need the following axioms:

Axiom of monotone sequence. For every monotone sequence of events $C_n \nearrow C$ and an event B such that

$$C_n \overset{\beta}{\leqslant} B, \text{ for all } n, \text{ then } C \overset{\beta}{\leqslant} B.$$

As Kraft, Pratt and Seidenberg (1959) showed, these axioms are still not sufficient to guarantee a probability measure; we need an additional axiom such as the following one in regard to the partitioning of events:

Axiom of partitioning of an event. Every event can be partitioned into two equally probable events.

Villegas (1964) showed under certain assumptions that this is equivalent to an axiom that there are no atoms, for example, Savage's axiom 6 is of this form.

Another approach to deriving the probabilities on β has been suggested by Anscombe and Aumann (1964). Their approach is closely related to the assumption that Θ contains a sufficient number of sets to extend the measure to β. We shall not consider this method here.

6.3 Empirical studies based on Axiom IV.

In this section we shall give the results of some empirical studies which are related to Axiom IV. First we present studies relating to the Ellsberg Paradoxes I and II which may contradict either Axiom II or Axiom IV though Paradox I at least seems to contradict Axiom IV rather than II.

<u>Empirical results of the type of Ellsberg Paradox I</u>. Most empirical studies have used real numbers as rewards (amounts of money which can be won) and the actions considered have reward functions of the following form:

$$X(\omega,a) = \begin{cases} \alpha r & \omega \in B \\ 0 & \omega \in \overline{B} \end{cases} \qquad X(\omega,c) = \begin{cases} \alpha r & \omega \in \overline{B} \\ 0 & \omega \in B \end{cases}$$

$$X(\omega,b) = \begin{cases} r & \omega \in D \\ 0 & \omega \in \overline{D} \end{cases} \qquad X(\omega,d) = \begin{cases} r & \omega \in \overline{D} \\ 0 & \omega \in D \end{cases}$$

where $D \in \Theta$ but $B \in \beta - \Theta$, and α any real number. If $\alpha = 1$ an "Ellsberg-type violation" is the choice $b \overset{A}{>} a$ and $d \overset{A}{>} c$.

<u>MacCrimmon (1965) study</u>. In a study with 38 business executives, MacCrimmon used a series of this type of problem where a mixture of "known" probabilities versus "unknown"

probabilities was used to determine if biases exist. The first problem considered was

B = {the stock price of Pierce Industries is increasing}

D = {a red card is drawn from a standard deck}

r = \$1,000 $\alpha = 1$.

In this type of problem the contradiction will only occur if the probability of the event B is close to 1/2. For example, if the economy is on the upswing and nearly all stocks have increased in recent days, it is not very likely that an "Ellsberg-type violation" will be obtained. The second type of problem was similar. In this case the events B and D were defined by {U.S. GNP increases next year} and {a coin landing heads} respectively.

In the first problem, 27 subjects were consistent with the axioms. That is, if they preferred action a over action b in the first set, then they preferred the action d over the action c in the second set. Seven subjects had violations of the event complement condition; five of these subjects had Ellsberg-type violations (i.e., they preferred the "known" card bet both times), while the other two preferred the stock bet both times. Hence the rate of violation was 21%. In the second problem there were 24 consistent subjects and 7 subjects who had Ellsberg-type violations. (The remaining 7 subjects had some degree of indifference.) Hence the rate of violation was 23%.

When the stakes were changed to yield \$10 more on the "unknown" event (i.e., $\alpha = 1.01$), the proportion of Ellsberg-type violations dropped to 12% (4 out of 34) on the first problem and 17% (6 out of 35) on the second problem. Note here, though, that the choice of the "unknown" stock bet in both sets cannot be called a violation with these payoffs since they pay more and would be the logical choice if the "known" and "unknown" events were deemed equally likely.

For the first problem, the subjects were also presented with reasons supporting consistent (i.e., axiom-based) responses and with reasons supporting Ellsberg-type violations. Nineteen subjects judged the consistent argument the more reasonable, while 12 subjects preferred the violating argument. Thus overall, there was a rate of 39% accepting the Ellsberg-type violation. Among those subjects who had consistent answers themselves, the rate of acceptance of the Ellsberg-type answer was 22% (6 out of 27), while among those who had an inconsistent answer themselves, the acceptance rate was 57% (4 out of 7).

<u>MacCrimmon and Larsson (1975) study</u>. In their study, 19 subjects were presented with two sets of 11 alternative wagers and were asked to rank the wagers in each set in order of their preference. The sets differed in terms of payoffs; in the first set r = \$1,000, $\alpha = 1$; in the second set, r = \$1,000, $\alpha = 1.01$. We shall only consider the part of their study which concerns the Ellsberg Paradox I.

$$R_{I} = \{\text{a red ball drawn from Urn I}\}$$

$$R_{II} = \{\text{a red ball drawn from Urn II}\}.$$

Fifteen of the 19 subjects had an Ellsberg-type violation. Another subject ranked three of the actions equal, with the fourth one less. Two subjects ranked all four actions equal, while the remaining subject ranked one of the actions highest and the other lowest with the known probabilities and the other ranked between them. Hence, only three of the 19 subjects behaved consistently with the utility axioms; there were 16 subjects with Ellsberg-type violations. This is a very high violation rate of 84%. Looking at it another way, 16 subjects preferred b to a, and three subjects were indifferent. Of these 16 subjects, only one preferred c to d (i.e., the bets on the complements were in the right order). They conclude from these results that bets on an urn with a specified composition seem to be preferable to bets on an urn with an unknown composition.

<u>Empirical studies of Ellsberg Paradox II</u>. In Ellsberg's paradox II we are concerned with three events, say B_1, B_2, and B_3, such that $B_i \cap B_j = \emptyset$ $i \neq j$ and $B_1 \cup B_2 \cup B_2 = \Omega$. We also assume that $B_1 \varepsilon \Theta$ but B_2, B_3 $\beta - \Theta$, and as in Ellsberg paradox I the rewards are "amounts of money which can be won". For simplicity we shall denote $\mu(B_1)$ by p. The actions which will be considered have reward functions defined by:

$$X(\omega,a) = \begin{cases} r & \omega \varepsilon B_1 \\ 0 & \omega \varepsilon \overline{B}_1 \end{cases} \qquad X(\omega,b) = \begin{cases} r & \omega \varepsilon B_2 \\ 0 & \omega \varepsilon \overline{B}_2 \end{cases}$$

$$X(\omega,c) = \begin{cases} r & \omega \varepsilon \overline{B}_2 \\ 0 & \omega \varepsilon B \end{cases} \qquad X(\omega,d) = \begin{cases} r & \omega \varepsilon \overline{B}_1 \\ 0 & \omega \varepsilon B_1 \end{cases}$$

We are then asked our preference between a and b and also between c and d. An Ellsberg-type violation would then be $a \overset{A}{>} b$ and $d > c$.

The Ellsberg problem can be described as: $p = 0.33$, B_1 = the event that (a red ball is drawn), B_2 = the event that (a black ball is drawn), B_3 = the event that (a yellow ball is drawn), and $r = \$1,000$. Since the number of balls in the urn may influence the decision we shall also denote this by n.

Slovic and Tversky (1975) study. They used the original Ellsberg problem ($p = 1/3$, B_1 = {red ball}, B_2 = {black ball}, B_3 = {yellow ball}, $n = 90$, and $r = \$1000$. Of their 29 college students, 19 made Ellsberg-type violations when the problem was first presented to them. On a second presentation of the same problem, there were 21 Ellsberg-type violations. In a second group of 49 subjects who were presented with arguments before making their choices, 38 subjects agreed with an Ellsberg-type of argument and 39 made an Ellsberg-type violation in their choices.

MacCrimmon and Larsson's study (1975). In their study, they varied the parameters p in the problem. This reflects the "known" chances of winning with event B_1 and correspondingly, there is a $1-p$ chance of winning with event "B_2 or B_3". They considered the following values of p: 0.20, 0.25, 0.30, $0.3\dot{3}$, 0.34, 0.40, and 0.50. One would expect the highest tendency to make Ellsberg-type violations around $p = 1/3$. To obtain some information on payoff levels they used $r = \$1,000$ and $r = \$1,000,000$. They used $n = 100$ balls in all cases except one presentation of the original Ellsberg problem with $n = 90$ balls (and $p = 1/3$) for direct comparison purposes.

In the original Ellsberg problem, 11 of the 19 subjects made the Ellsberg-type violation (i.e., a,d). Only five individuals made choices conforming to the axioms (i.e., a,c or b,d). Hence there does seem to be a high rate of violation of the notion that probabilities can be assigned to "uncertain" events. In the form of the problem we used (with n = 100 balls), the rate of violation tended to be even higher. For p = 0.33, or 0.34, 70% of the subjects made Ellsberg-type violations. So there was considerable violation for these particular parārameters.

In an urn with 100 balls with proportion p of red balls, we would expect that when p is close to 0, a choice of b,d and when p is close to 1, would expect a choice of a and c. However, when p is around 1/3, individuals may be somewhat indifferent about the choices and may then choose on the basis of how well the chances are known. Since the chances for alternatives a and d are specified, we would expect a much higher proportion of such violations around this value of p. As p deviates from 1/3, then although the chances are still specified with alternatives a and b, one is accepting a rather low chance of winning by taking a when p is small or d when p is large.

Regardless of whether this is the rationale for the choices, they definitely observed this kind of behaviour. It was observed that 70% of the choices for p = 0.33 or 0.34 are inconsistent with the axioms, but for p = 0.20, this percentage drops to 14% and for p = 0.50 it drops to 0%.

Another study directly related to Axiom IV was made by E. M. Shuford (1959) who designed and constructed a set of 20 x 20 matrices of small lines which would either be in a vertical or horizontal position. The subjects studied these matrices for a short period of time and were then asked what percentage of lines were vertical. Here the exact number of vertical lines was known to the experimenter, and therefore a study of how accurate the subjects were in determining the fraction of vertical lines could be determined. Shuford found that small percentages were nearly always overestimated and larger percentages underestimated which would suggest the following graphical representation between W and μ.

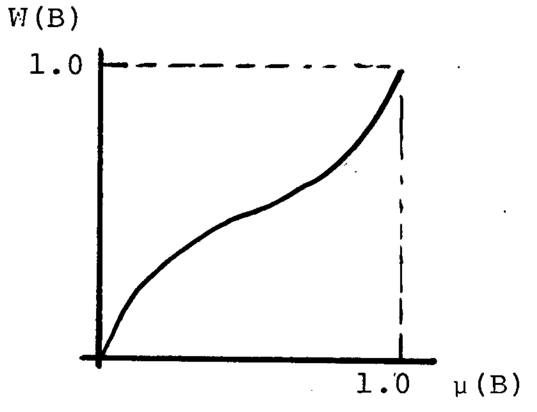


Fig. 6.3. Relation between μ(B) and W(B) by experiment of E. M. Shuford

This does not suggest that W ought not to be a non-linear function of μ. It only illustrates the difficulties of obtaining correct probabilities.

6.4 Relation to other axiom systems

Marschak's, vonNeumann & Morgenstern's and Arrow's approaches assume a probability measure as we have given for all sets in β. Their approaches to the expected utility theory would therefore have little interest to section 6. Savage's and Luce & Krantz's approaches are slightly different. In their approaches an ordering is induced on all possible events in such a way that a probability measure may be specified on those events. That is, the ordering induced on the events must satisfy the conditions specified in section 6.2. Savage's approach does not satisfy the axiom of monotone sequence, and can, therefore, only specify a finite additive measure.

6.5 Alternative to Axiom IV

An alternative approach would be to remove the condition that W behave strictly as a measure. For example, Fellner (1961) found that in some cases when he tried to derive subjective probabilities that $W(B) + W(\bar{B}) = \gamma$ where $\gamma \neq 1$. If $W(B) = k\mu(B)$ $(\mu > 0)$ it would not be difficult to derive a theory similar to the expected utility theory. However, he also found that k is not a constant but is a function of B. Hence Axiom II would also be contradicted. We shall discuss his approach in addition to a new approach in Part II.

7.0 Summary of Part I

In this section we shall show that the axioms we have considered here are both necessary and sufficient for the expected utility criterion to hold. Before doing so, however, we shall summarize the basic assumptions and axioms made.

<u>Summary of the assumptions</u>

In Assumption 1 we assumed $X(\cdot,a)$ was a measurable function with respect to Ω. Axiom III assumes the existence of a measurable function U on R, and hence we may now consider the composite function $UX(\cdot,a)$ for all a. The most important assumption concerning measurability is that $UX(\cdot,a)$ is a measurable function with regard to the Borel sets on the real line and β. This follows from the fact that if U and X are both measurable then UX is measurable (Halmos, 1950, Theorem B, pp.162).

<u>Assumption 1</u>. There exist:

i) a probability space (Ω,Θ,μ), where Ω is the set of states,

ii) a measurable space (R,Ψ), where R is the set of rewards or outcomes,

iii) an index set A, called an action space, such that for each $a \in A$ there exists a reward function $X(\cdot,a)$ from Ω to R,

iv) a σ-algebra β of Ω such that $\Theta \subset \beta$ and each function $X(\cdot,a)$ is β-measurable, and

v) a relation $\overset{A}{\leqslant}$ on A.

The following axioms were also made with respect to the decision maker's behaviour.

Axiom I. Existence Axiom

There exists a real-valued function f on A such that for any a,b ε A,

$$\text{if } a \overset{A}{\leqslant} b \text{ then } f(a) \leqslant f(a), \text{ and}$$

$$\text{if } a \overset{A}{<} b \text{ then } f(a) < f(b).$$

From a practical viewpoint this axiom suggests that there exists a monetary amount which we would be willing to pay or receive in order to obtain or sell each alternative. It does not suggest the amount we are willing to pay but only the existence of an amount. If there are alternatives which are "above rubies", or impossible to specify in monetary rewards, this axiom would not have a meaning and would not be acceptable.

Axiom II. Additivity Axiom

There exists a real-valued function h on β x A such that

a) $f(a) = h(\Omega,a)$

b) for $\{B_i\}$ i=1,... such that $B_i \ \varepsilon \ \beta$ for all i and $B_i \cap B_j = \emptyset$ for $i \neq j$ then $h(\bigcup B_i, a) = \sum_{i=1}^{\infty} h(B_i, a)$ for all a ε A

c) for any B ε β, and for any a,b ε A such that $X_B(\cdot,a) = X_B(\cdot,b)$, then $h(B,a) = h(B,b)$.

Axiom II implies that the events are independent of each other. However, we do not mean a statistical independence, but that the evaluation of an action by considering the reward for a given event is not affected by the possible rewards on the complement of that event. Therefore, we can not argue that we are willing to take a small chance on a given event because we have a relatively conservative reward elsewhere.

Recall that in Axiom IIIb and c we use the notation $h(B,z)$ defined by $\sum_{i=1}^{k} W(B_i)\alpha_i$ where z identifies the simple function $Y(\omega,z) = \alpha_i, \omega \in B_i$ for $i=1,\ldots,n$ and $\bigcup_{i=1}^{k} B_i = B$ (see section 5.1).

<u>Axiom III</u>. Separability Axiom

a) There exists a non-negative real valued σ-additive function W on β and a real valued measurable function U on R such that for any $a \in A$, $B \in \beta$ and $r \in R$, if $X_B(\cdot,a) = r$, then

$$h(B,a) = W(B)U(r).$$

b) Let r_o be any fixed reward in R, and let $X(\cdot,b)$ be any reward function such that $X_B(\omega,b) \overset{R}{\geqslant} r_o$ for all $\omega \in B$, for $B \in \beta$. If Z_o is the set of all simple functions z such that $Y_B(\cdot,z) \leqslant UX_B(\cdot,b)$, then $h(B,b)$ satisfies

$$h(B,b) = \sup_{z \in Z_o} h(B,z)$$

c) Similarly for any reward function $X(\cdot,c)$ such that $r_o \overset{R}{>} X_B(\omega,c)$ for all $\omega \in B$, if Z_1 is the set of all simple functions z such that $Y_B(\cdot,z) \geqslant UX_B(\cdot,c)$, then $h(B,c)$ satisfies

$$h(B,c) = \inf_{z \varepsilon Z_1} h(B,z)$$

Axiom IV. For any $B \varepsilon \Theta$, $W(B) = \mu(B)$.

Axioms III and IV specify an independence between the reward and the probability of receiving this reward. This implies that we have equal satisfaction from winning $1,000 on the stock exchange or from winning a lottery ticket. This may not always be true, of course, since in the first case we may feel a given satisfaction because receiving $1,000 from buying and selling stock involves skill, but with a lottery ticket, only luck would determine the winner. Of course, in this case the reward set may be modified so as to include more than a monetary reward. In this case, however, it becomes very difficult to determine any utility function.

These axioms and assumptions are necessary and sufficient for the existence of a function V on R such that

$$a \overset{A}{\geqslant} b \text{ implies } \int VX(\cdot,a)dW \geqslant \int VX(\cdot,b)dW,$$

and

$$a > b \text{ implies } \int VX(\cdot,a)dW > \int VX(\cdot,b)dW.$$

To show this it is sufficient by Axiom I to show that

$$f(a) = \int VX(\cdot,a)dW.$$

<u>Theorem 7.1</u> If Axioms I, II, III and IV hold then there exists a real-valued measurable function V on R such that

$$f(a) = \int VX(\omega,a)dW$$

<u>Proof</u>. By Axiom III there exists a real-valued function U on R, so define $V(r) = U(r)$.

If $a \varepsilon A$ and $X(\omega,a) = r$ for all $\omega\varepsilon\Omega$ then by Axioms II, III and IV respectively

$$f(a) = h(\Omega,a) = W(\Omega)U(r) = U(r) = V(r)$$

and

$$f(a) = \int UX(\omega,a)dW = \int VX(\omega,a)dW.$$

Therefore, the theorem holds for constant functions.

Suppose instead that $a \varepsilon A$ and $X(\cdot,a)$ is a simple function, that is, there exists a partition $B_1,\ldots,B_n$ of Ω such that $X_{B_i}(\cdot,a) = r_i$ $i = 1,\ldots,n$. Then Axioms IIa, III and the definition of the integral imply

$$f(a) = \sum_{i=1}^{n} h(B_i,a)$$

$$= \sum_{i=1}^{n} W(B_i)\ U\ (r_i)$$

$$= \int UX(\omega,a)dW$$

$$= \int VX(\omega,a)dW.$$

Lastly, consider $a \varepsilon A$ for an arbitrary reward function $X(\cdot,a)$. Let r_o be any fixed reward in R. Axioms IIa, IIIb and IIIc respectively guarantee that

$$f(a) = h(\{X(\cdot,a) > r_o\},a) + h(\{X(\cdot,a) < r_o\},a)$$

$$= \sup_{z \varepsilon Z_o} h(\{X(\cdot,z) \geqslant r_o\},z) + \inf_{z \varepsilon Z_1} h(\{X(\cdot,z) < r_o\},z)$$

where Z_o is the set of all simple functions that $Y(\omega,z) \leqslant UX(\omega,a)$ for all $\omega \varepsilon \{X(\cdot,a) \geqslant r_o\}$ and Z_1 is is the set of all simple functions such that $Y(\omega,z) \geqslant UX(\omega,a)$ for all $\omega \varepsilon \{X(\cdot,a) < r_o\}$.

Hence $h(\{X(\cdot,a) > r_o\},a)$ is equal to the supremum of the integral of all simple functions less than $UX(\cdot,a)$, similarly for $h(\{X(\cdot,a) < r_o\},a)$. This is the definition of the integral $\int UX(\cdot,a)dW$ and completes the proof.

What is also true is, of course, that if f is to be expressed as the expected utility, the axioms must also hold.

Theorem 7.2. Given the objects in Assumption 1, suppose that there exists a measure ν on β such that $\nu(B) = \mu(B)$ for all $B \varepsilon \Theta$, and a real-valued measurable function V on R such that if a > b then

$$\int VX(\omega,a)\ d\nu > \int VX(\omega,b)\ d\nu$$

and if $a \geqslant b$ then

$$\int VX(\omega,a)\,d\nu \geqslant \int VX(\omega,b)\,d\nu.$$

Then Axioms I, II, III, and IV are satisfied with $W = \nu$, $U(\cdot) = V(\cdot)$ and

$$f(\cdot) = \int UX(\omega,\cdot)\,dW.$$

<u>Proof</u>. Define $f(a) = \int VX(\omega,a)\,dW$, for all $a \in A$. Then Axiom I is obviously satisfied.

Let B_i $i=1,2,\ldots$ form a partition of Ω such that $B_i \in \beta$ for all i, then for any $a \in A$

$$f(a) = \int VX(\omega,a)\,d\nu = \sum_i \int_{B_i} VX(\omega,a)\,d\nu.$$

Define $h(B_i,a) = \int_{B_i} VX(\omega,a)\,d\nu$, then clearly $f(a) = \sum_i h(B_i,a)$.

Hence Axiom IIa must hold.

If $X_B(\cdot,a) = X_B(\cdot,b)$ then $VX(\cdot,a) = VX(\cdot,b)$ for all $\omega \in B$. Hence

$$\int_B VX(\omega,a)\,d\nu = \int_B VX(\omega,b)\,d\nu \text{ or } h(B,a) = h(B,b).$$

Hence Axiom IIb must hold.

If for any $B \in \beta$, $X_B(\cdot,a) = r$ for some $a \in A$ then

$$h(B,a) = \int_B VX(\omega,a)\,d\nu$$

$$= V(r)\,\nu(B)$$

Define $U(r) = V(r)$ and $\nu(B) = W(B)$ then Axiom IIIa is satisfied. Axioms IIIb and IIIc then follow from the definition of the integral.

Axiom IV follows from the assumption that $\mu(B) = \nu(B) = W(B)$ for $B \varepsilon \Theta$.

We have, therefore, shown that the axioms specified are both necessary and sufficient for the expected utility theory to hold. Thus, the theory of maximization of the expected utility can only be accepted if we agree with each of the axioms specified here, and their implications. Hopefully, we have illustrated the main difficulties involved with each of the axioms, giving a greater understanding for their acceptance or rejection.

The two most difficult axioms to accept are Axiom II and Axiom IV. We have, therefore, spent the most time on them to illustrate their implications and relations to other axioms -- at least for Axiom II. Axiom IV will be considered in much more detail in Part II of the thesis. We shall also consider Axiom II there, and suggest an alternative to the additivity axiom.

1.0 Introduction to Part II

Decision making under partial risk

Classical decision making is usually categorized in accordance with the decision maker's knowledge of the consequences (rewards) of his alternatives (actions). Specifically, decision problems are classified as being under certainty, risk and uncertainty (Knight, 1921). Certainty means that one reward is specified for each of the decision maker's alternatives. In the case of risk, the probability of the consequences is known for all consequences and for all actions. Uncertainty means that only the set of possible consequences is known.

Evaluation of the decision maker's alternatives

We shall assume as before that there exists an evaluation function on the set of actions; that is, a real-valued function $f(\cdot)$ exists on A such that if one action is preferred to another, the numerical value associated with the first action is greater than that of the second. It is also assumed that there exists a complete ordering on the rewards which can be represented by a real-valued order-preserving function. We shall call this function a utility function.

The utility function is easier to determine than the evaluation function since the relationship between actions and consequences implies that the total number of possible utility functions is a subset of the total number of possible evaluation functions.

In decision theory under certainty, each action gives only one possible reward and we assume that the evaluation function on an action takes the same value as the utility function for the corresponding reward. In this case the specification of a utility function is equivalent to the specification of the evaluation function.

In decision making under either risk or uncertainty it is often assumed that the utility function can easily be specified (or derived by some rule(s)) and the evaluation function is then specified in terms of the utility function.

For example, in decision making under risk several evaluation functions have been suggested, although the most accepted is the expected value of the utility function (Bernoulli, 1738). In Part I we considered the evaluation function defined by expected utility in some detail. Under uncertainty there has not been an evaluation function specified which has been uniformly accepted. Five functions are frequently suggested (although more exist). We shall briefly state them here. They all assume the existence of a utility function, and the evaluation function is specified in terms of this utility function.

<u>The maximax criterion (Hurwitz, 1951)</u>. The evaluation function for each action is specified by taking the maximum utility over all possible rewards of that action.

<u>The maximin criterion (Wald, 1950)</u>. The evaluation function for each action is specified by taking the minimum utility of all possible rewards for each consequence of that action.

The Hurwitz α-criterion (Hurwitz, 1951). The evaluation function is specified by a linear combination of the maximum and minimum utility for each action.

The principle of insufficient reason (Bernoulli, 1738). The evaluation function is specified as the average utility for each action.

The expected utility criterion (Savage, 1954). A probability distribution is "derived" on the consequences. The evaluation function is then equal to the expected utility for each alternative.

For a critical review of each of these evaluation functions, see Milnor (1954).

Partial Risk Problems

A fourth category of decision problems can be specified as falling between risk and uncertainty; that is, some knowledge of the probabilities of rewards are known, but a probability distribution can not be completely specified. This category has been recognized by decision theorists for some time and is commented on in standard textbooks in the area. For example,

> "A common criticism of such criteria as the maximin utility, minimax regret, Hurwitz-α, and that based on the principle of insufficient reason is that they are rationalized on some notion of complete ignorance. In practice, however, the decision maker usually has some vague partial information concerning the true state. No matter how vague it is, he may not wish to endorse any characterization of complete ignorance, and so the heart is cut out of criteria based on this notion."
>
> (Luce and Raiffa, 1957, p.299)

Knight also asserts that such overall judgments may influence decision:

> "The action which follows upon an opinion depends as much upon the amount of confidence in that opinion as it does upon the favorableness of the opinion itself ...Fidelity to the actual psychology of the situation requires we must insist recognition of these two separate exercises of judgment, the formation of an estimate and the estimation of its value."
>
> (Knight, 1921, p.227)

Decision problems which fall into this category are called partial risk problems (other common names are decision making under partial ignorance, or partial uncertainty). A partial risk problem may therefore be very close to a risk problem if the probability distribution can nearly be specified and similarly it may be very close to an uncertainty problem if very little can be specified of the probability distribution. We can therefore consider risk and uncertainty as extreme cases of partial risk. It is not surprising therefore to find that the existing methods for specifying an evaluation function for a partial risk problem fall between those used in evaluation the risk and uncertainty problems.

Evaluation of partial risk problems

There are basically three methods used in describing the evaluation function on alternatives. They can be classified as follows:

1) Translating the problem into a risk problem. That is, assuming the existence of a probability distribution over the reward the evaluation function is then specified as in the case of decision making under uncertainty.

2) Combining the evaluation functions used for risk and uncertainty.

3) Deriving a "preference function" on the uncertain rather than a probability measure.

We briefly discuss each method here to illustrate the differences from the approach we shall discuss in this thesis.

Evaluation Method 1. The exact probability distribution is assumed to exist but is not necessarily known. A probability measure has been specified in two related methods. The first of these is a derivation of subjective probabilities. A probability distribution is derived from the preference among alternatives. This method has been used by Ramsey (1926), De Finetti (1937), and Savage (1954). The subjective probability measure usually has the same properties as Kolmogorov's probability axiom (1933), except that sometimes finite additivity is assumed rather than σ-additivity.

The second method used to derive a probability is to assume a set of weights on the set of possible probability distributions and to use the combined probability measure to calculate the expected utility. This method is sometimes used in statistics (see, for example, Good, 1965).

Once the probability measure is specified the problem has been reduced to a decision problem under risk. Hence any evaluation function used for decision making under risk can also be used in these cases. The most common evaluation function is the expected value of the utility function. This approach, therefore, does not differentiate between risk and uncertainty. Savage recognized that this approach might create difficulties:

> "...there seem to be some probability relations about which we feel relatively "sure" as compared with others...The notion of "sure" and "unsure" introduced here is vague, and my complaint is precisely that neither the theory of personal probability, as it is developed in this book, nor any other device known to me renders the notion less vague..."
>
> (Savage, 1954, pp.57-58,59)

Evaluation method 2. The second evaluation method combines the evaluation functions used for risk and uncertainty. There are also two basic approaches for this method. Both of these approaches assume that a set of possible probability distributions can be specified on the consequences. The first (see Good, 1965) assigns a set of weights on the possible probability distributions and the expected value is calculated using the combined probability distribution. The evaluation function is specified as a linear combination of any of the criterion under uncertainty and the expected utility. There has only been one suggested so far and that is a linear combination with weights greater than zero, and sum to one between expected utility and the maximin (see Ellsberg, 1961). It is obvious that his approach could also have been used for any of the other evaluation functions for decision making under uncertainty. The coefficient of the linear combination would depend on how "close" the problem would be to a decision problem under risk versus uncertainty. We shall discuss this method in greater detail in section 6.1.

The second approach has been to combine maximin with the expected utility theory in the following way. For each action a set of possible distributionson the rewards is determined, then the evaluation function is specified as the minimum of all possible expected values for a given action. This approach has been the most common (see, for example, Menges, 1966; Blum and Rosenblatt, 1968; Randles and Hollander, 1971).

Evaluation method 3. The last of the three methods of evaluation is similar to Savage's in the sense that a function is derived on the reward space indicating in some way the likelihood of receiving the rewards. In this case, however, this function may not be a probability measure. The justification for this would be that it is easier to determine a probability distribution of rewards for some actions than for others. This is therefore an extension of the Savage approach in the sense that this method recognizes the differences between "sure" and "unsure" events. This function will be called a preference function rather than a probability measure.

Once the preference function has been specified, a difficulty arises from how to use this to specify the evaluation function. Fellner (1961) suggested one method of doing this. He assumed that the preference function must have the properties that a probability measure on the rewards can be derived from it. The expected utility was then calculated using the derived probability measure. However, in his approach, the utility function is different from the utility function under risk since the utility under partial risk contains an element of gambling.

In Part II of the thesis we shall suggest an alternative method of evaluation partial risk problems. Of those methods so far described this new method is closest to Fellner's approach. There are some fundamental differences, however. Following Fellner, a preference function is first derived on the possible rewards for each action. Here, however, we do not assume this preference can be transformed into a probability measure on the sets. We also specify a set of axioms which we believe the decision maker ought to follow from which an evaluation function may be specified.

In section 2 we shall specify the underlying assumption axioms and notation we shall use. We shall also specify the problem in more mathematical terms.

In the theory developed here we shall assume three basic axioms for problems under partial risk. These are stated and discussed in section 3, 4 and 5 respectively. In section 6 we shall summarize all the axioms and assumptions for easy reference, and also show that an evaluation function is completely specified given these axioms and assumptions.

In section 7 we shall state an additional axiom that simplifies the practical aspects of partial risk problems but which is not necessary for the theory. We shall also suggest some other simplifications. In section 8 we shall consider the preference functions we have derived as a probability measure. In section 9 we shall give some support for the theory from some empirical studies which have been made, and finally in section 10 we shall summarize the results.

2.0 The basic assumptions

In this section we shall specify the notation and some of the basic assumptions needed for the model developed here. Most of the notation follows that used in Part I. We shall assume that there exists a probability space (Ω,Θ,μ), a reward space (R,Ψ) and an action set, A, with a complete ordering $\overset{A}{\geqslant}$ on A. For the time being we shall let A be an arbitrary set and $\overset{A}{\geqslant}$ an arbitrary complete ordering and restrict the set A and the ordering $\overset{A}{\geqslant}$ as needed in subsequent sections. For a given action $a \in A$, the reward function is a function from Ω to R, and is denoted by $X(\cdot,a)$. The set of all Θ-measurable reward functions in A will be indexed by the subset A_o of A, that is if $a \in A_o$ then $X(\cdot,a)$ is measurable (we shall modify this definition of A_o presently). The smallest σ-algebra of subsets of Ω, such that all functions $X(\cdot,a)$, $a \in A$, are measurable will be denoted by β. Necessarily, then, $\Theta \subset \beta$. (The case of $\Theta = \beta$ will not be considered.)

With this notation we can summarize the standard categories of decision making as follows:

A given action $a \in A$, is a decision under

1) <u>certainty</u>, if $a \in A$ and there exists a $B \in \Theta$ such that $X_B(\omega,a) = r$ and $\mu(B) = 1$,
2) <u>risk</u>, if $a \in A_o$,
3) <u>uncertainty</u>, if $a \in A$, and if for any set $B \subset \Omega$ such that $B \neq \emptyset$, $B \neq \Omega$ and $B = \{\omega : X(\omega,a) \in C \text{ for some } C \in \Psi\}$, then $B \notin \Theta$,
4) <u>partial risk</u>, if a is any alternative in A.

In these definitions the categories 1), 2), and 3) are all included in 4). This classification is, therefore, redundant if an acceptable criterion may be found for partial risk problems. Our aim is to specify a criterion for partial risk problems. The same criterion may also, therefore, be used for uncertainty, risk, or certainty problems.

In section 2.1 we shall state three fundamental axioms which are the basis for our development, and in section 2.2 we shall illustrate the difference between β and Θ, from a practical decision making approach.

2.1 Statement of the axioms

In this section we shall specify three axioms and two assumptions which jointly guarantee the existence of a utility function U on R, and also specify the expected utility criterion for actions belonging to A_o.

The first axiom we shall state here is similar to Axiom I in Part I. That is, it assumes the existence of a function on A, which preserves the ordering on A.

<u>Axiom I</u>. There exists a real-valued function $f(\cdot)$ on A such that

$$a \overset{A}{\geqslant} b \text{ if and only if } f(a) \geqslant f(b)$$

The objections and alternatives to this axiom were discussed in Part I, and will, therefore not be discussed here. Other axioms which we shall assume impose conditions on the ordering $\overset{A}{\geqslant}$ on A and the membership of A, or equivalently, they specify properties of the function f.

If for all $r \in R$ there exists an $a \in A_o$ such that $X(\omega,a) = r$ for all $\omega \in \Omega$, then, in general, a function U may be defined on R by $U(r) = f(a)$. A difficulty arises if there also exists a $b \in A_o$ such that $X(\omega,b) = X(\omega,a)$ for all $\omega \in \Omega$ but $f(b) \neq f(a)$. In this case U is not uniquely defined. Therefore, for U to be a uniquely defined function we need both the existence of all constant functions (Assumption 1) and also the assumption that if two constant reward functions are equal they have the same preference (Axiom IIi).

<u>Assumption 1</u>. For each $r \in R$, there exists an $a \in A_o$ such that $X(\omega,a) = r$ for all $\omega \in \Omega$.

Jointly Axiom IIi and Assumption 1 also imply that an ordering may be induced on R, since this is a weaker condition than the existence of U. The ordering we shall consider on R is defined as follows:

If $X(\omega,a) = s$ and $X(\omega,b) = r$ for all $\omega \in \Omega$ then

$$r \overset{R}{\geqslant} s \text{ if } f(a) \geqslant f(b),$$

or equivalently

$$r \overset{R}{\geqslant} s \text{ if } U(r) \geqslant U(s).$$

For any ordering $\overset{R}{\geqslant}$ on R, dominance can be defined among the reward functions in the usual way: the reward function $X(\cdot,a)$ <u>dominates</u> the reward function $X(\cdot,b)$ with respect to $\overset{R}{>}$ if

$$X(\omega,a) \overset{R}{\geqslant} X(\omega,b) \quad \text{for all } \omega \in \Omega$$

(or reward function $X(\cdot,b)$ is dominated by reward function $X(\cdot,a)$). That is, we prefer the reward of $X(\omega,a)$ to the reward of $X(\omega,b)$ for all possible outcomes. It seems reasonable to assume that if $X(\cdot,a)$ dominates $X(\cdot,b)$, $a,b \in A$ and the ordering on R has been induced by the constant reward functions, then $f(a) \geqslant f(b)$.

In the further development we shall assume that when we discuss an ordering on R or dominance among reward functions we shall assume that the ordering has been induced by the constant functions. We are now ready to state Axiom II completely.

<u>Axiom II</u>. i) If for any $r \in R$ there exists $a,b \in A$ such that $X(\omega,a) = r$ and $X(\omega,b) = r$ for all $\omega \in \Omega$, then $f(a) = f(b)$.

ii) If $X(\cdot,a)$ dominates $X(\cdot,b)$ then $a \overset{A}{\geqslant} b$.

A natural extension of Axiom IIi would be to assume that if $X(\omega,a) = X(\omega,b)$ for all $\omega \in \Omega$ then $f(a) = f(b)$. This follows, however, directly from part ii of the Axiom as follows:

If $X(\cdot,a) = X(\cdot,b)$ then $X(\omega,a) \overset{R}{\geqslant} X(\omega,b)$ for all $\omega \in \Omega$ and hence $X(\cdot,a)$ dominates $X(\cdot,b)$ by Axiom IIii $f(a) \geqslant f(b)$. Similarly $X(\cdot,b)$ dominates $X(\cdot,a)$ and hence $f(b) \geqslant f(a)$ and therefore $f(a) = f(b)$.

If $a \in A_o$ and $X(\cdot,a)$ is a constant reward function then U is defined by the equality $UX(\cdot,a) \equiv f(a)$. The function U will

$$X(\omega,a) \overset{R}{\geq} X(\omega,b) \quad \text{for all } \omega \in \Omega$$

(or reward function $X(\cdot,b)$ is dominated by reward function $X(\cdot,a)$). That is, we prefer the reward of $X(\omega,a)$ to the reward of $X(\omega,b)$ for all possible outcomes. It seems reasonable to assume that if $X(\cdot,a)$ dominates $X(\cdot,b)$, $a,b \in A$ and the ordering on R has been induced by the constant reward functions, then $f(a) \geq f(b)$.

In the further development we shall assume that when we discuss an ordering on R or dominance among reward functions we shall assume that the ordering has been induced by the constant functions. We are now ready to state Axiom II completely.

<u>Axiom II</u>. i) If for any $r \in R$ there exists $a,b \in A$ such that $X(\omega,a) = r$ and $X(\omega,b) = r$ for all $\omega \in \Omega$, then $f(a) = f(b)$.

ii) If $X(\cdot,a)$ dominates $X(\cdot,b)$ then $a \overset{A}{\geq} b$.

A natural extension of Axiom IIi would be to assume that if $X(\omega,a) = X(\omega,b)$ for all $\omega \in \Omega$ then $f(a) = f(b)$. This follows, however, directly from part ii of the Axiom as follows:

If $X(\cdot,a) = X(\cdot,b)$ then $X(\omega,a) \overset{R}{\geq} X(\omega,b)$ for all $\omega \in \Omega$ and hence $X(\cdot,a)$ dominates $X(\cdot,b)$ by Axiom IIii $f(a) \geq f(b)$. Similarly $X(\cdot,b)$ dominates $X(\cdot,a)$ and hence $f(b) \geq f(a)$ and therefore $f(a) = f(b)$.

If $a \in A_o$ and $X(\cdot,a)$ is a constant reward function then U is defined by the equality $UX(\cdot,a) \equiv f(a)$. The function U will and is made for convenience. Since the integral always exists for bounded measurable functions $\int UX(\omega,a)d\mu$ must exist for all $a \in A_o$.

We shall use the notation

$$EUX(\cdot,a) \quad \text{for} \int_{\Omega} UX(\omega,a)d\mu.$$

Axiom III. If $b \in A_o$ then $f(b) = EUX(\cdot,b)$.

To summarize, Assumption 1, Axiom I and Axiom IIi guarantee the existence of a utility function U on R, and also an ordering $\overset{R}{\leq}$ on R. Assumption 2 implies that $\int UX(\cdot,a)d\mu$ is well defined for all $a \in A_o$ and Axiom III that the expected utility criterion is the accepted criterion for decision making under risk. The axioms which follow in sections 3-6 will extend this evaluation function from A_o to A.

2.2 Comments on the assumptions

We have assumed the existence of a probability space (Ω,Θ,μ); that is, there exist events (members of Θ) for which the probabilities are known or at least the decision makers are willing to accept certain events for which they believe they know the probabilities. Consider, for example, the following events:

B_1 = {heads occurs when a coin is tossed}

B_2 = {a five occurs when a die is rolled}

B_3 = {Dow-Jones will close higher tomorrow than today}

B_4 = {the temperature tomorrow will reach a maximum value of 10°C}.

Some of these may be classified as belonging to Θ and others as belonging to β. Many people would be willing to accept $\mu(B_1) = 1/2$ and $\mu(B_2) = 1/6$. The probabilities for the events B_3 and B_4 are such that we may be less willing to specify them exactly. (The assumptions do not prevent us from doing so of course. For example a meteorologist may feel confident about specifying an exact probability for event B_4.) The idea that some probabilities can be "accurately" specified and others not, has been suggested by Anscombe and Aumann (1963). They coined the phrase "horse lottery" for those events for which probabilities may not be completely specified. Those events where a probability can be specified they called roulette lotteries. Fellner (1961) also separated events for which the probability may be stated with some accuracy and those for which this is not possible. He suggested that sets which belong to Θ are those for which the subjective probability may be supported by frequency probabilities as, for example, the drawing of cards from a deck of guaranteed composition.

Savage (1954) included all subsets of Ω in Θ. Hence he would hold that all probabilities can be specified exactly. Good (1965) disagrees that the probabilities can be exactly determined. He writes

> My own view, following Keynes and Koopman, is that judgments of probability inequalities are possible but not judgments of exact probabilities; therefore a Bayesian should have upper and lower betting probabilities.
>
> (Good, 1965; pp 5)

In the approach developed here we agree with Anscombe and Aumann (1963) that there exist both "horse lotteries" and "roulette lotteries". We also accept in part Good's argument that probabilities may not necessarily be exactly determined, although we allow it to be specified as precisely as one chooses by making a judgment of a sufficient number of probability inequalities.

3.0 The P-measure axiom

In section 2 the expected utility criterion was suggested for actions belonging to A_o. In extending this criterion in the usual mathematical fashion to all of A, two difficulties arise which are interrelated. One difficulty is that the required extension of the measure μ to the σ-algebra β may not exist. A property such as σ-additivity of the measure must often be sacrificed. However, when such a property is not satisfied by the measure it is difficult to define expected value.

In this section we shall assume the existence of a set function on β which is an extension of the measure μ. The existence of such a function will depend on what properties it is assumed to have. The properties we shall specify in the following section, however, will not put any restriction on β.

In section 3.1 we shall specify the first property of the extended measure. In section 3.2 we shall discuss some of the more important implications of this property and also give an example illustrating the assumptions so far made. Finally, in section 3.3 additional assumptions will be made to enable us to determine the values of the function for each $B \varepsilon \beta$.

3.1 Statement of the axiom

The concept of Axiom IV can be illustrated by the following example. Consider two reward functions defined by

$$X(\omega,a) = \begin{cases} r & \omega \in B \\ s & \omega \in \overline{B} \end{cases} \quad \text{and} \quad X(\omega,b) = \begin{cases} r & \omega \in D \\ s & \omega \in \overline{D} \end{cases}$$

where $B,D \in \Theta$, and $U(r) > U(s)$.

From Axiom III it follows that:

$$f(a) = U(r)\mu(B) + U(s)\mu(\overline{B}) \text{ and } f(b) = U(r)\mu(D) + U(s)\mu(\overline{D})$$

and hence

$$f(a) - f(b) = [U(r) - U(s)][\mu(B) - \mu(D)].$$

Therefore,

$$f(a) \geqslant f(b) \text{ if and only if } \mu(B) \geqslant \mu(D),$$

so that the preference between a and b can be determined by comparing the probabilities $\mu(B)$ to $\mu(D)$. Thus, for alternatives with the same two possible rewards the alternative with the largest probability of receiving the higher of the two rewards is chosen.

If P is the extended set function of μ to β, it seems reasonable to assume that P has the same property. Therefore if P is assumed to be a probability measure the assumption is clearly consistent with the expected utility criterion. This assumption is formalized in Axiom IV. Since the extension may not satisfy the requirement of what is generally called a probability measure (see definition in Appendix II, page 235), we shall call the extension a preference measure, or simply a P-measure.

We shall first assume that all reward functions with only two rewards belong to A.

<u>Assumption 3</u>. For any $D \varepsilon \beta$ and for any $r,s \varepsilon R$ there exists $b \varepsilon A$ such that

$$X(\omega,b) = \begin{cases} r & \omega \varepsilon D \\ s & \omega \varepsilon \overline{D} \end{cases} .$$

<u>Axiom IV</u>. There exists a set function P defined on β such that

a) if $B \varepsilon \Theta$ then $P(B) = \mu(B)$,

b) for any $r,s \varepsilon R$ such that $U(r) > U(s)$, and for any $B,D \varepsilon \beta$, let a,b be elements of A for which the reward functions are

$$X(\omega,a) = \begin{cases} r & \omega \varepsilon B \\ s & \omega \varepsilon \overline{B} \end{cases} \quad \text{and} \quad X(\omega,b) = \begin{cases} r & \omega \varepsilon D \\ s & \omega \varepsilon \overline{D} \end{cases} .$$

Then $a \overset{A}{\geqslant} b$ if and only if $P(B) \geqslant P(D)$.

We showed one implication of the expected utility criterion to be that if two reward functions are given by

$$X(\omega,c) = \begin{cases} r & \omega \in \overline{B} \\ s & \omega \in B \end{cases} \quad \text{and} \quad X(\omega,d) = \begin{cases} r & \omega \in \overline{D} \\ s & \omega \in D \end{cases}$$

with $B,D \in \Theta$, then $d \overset{A}{>} c$, when $a \overset{A}{>} b$. This follows since $a \overset{A}{>} b$ implies the probability of B is greater than the probability of D, so that the probability of $\overline{D}$ must be greater than the probability of $\overline{B}$ (since $\mu(B) + \mu(\overline{B}) = 1$ if $B \in \Theta$) and hence $d \overset{A}{>} c$.

If, however, the probabilities are not known, then the preference $a \overset{A}{>} b$ does not imply that the probability of B is greater than the probability of D, it only implies for some reason that we prefer alternative a to alternative b. For example, alternative a may seem less risky in the sense that we have some information of the likelihood of B occurring, and no information of the likelihood of D occurring. That is, we may differentiate between "known" versus "unknown" probabilities.

Axiom IV gives us a method to determine a partial ordering for all actions which would result in one of two rewards. If $a \overset{A}{>} b$, with a reward function as given in Axiom IV, then for actions e and f with reward functions

$$X(\omega,e) = \begin{cases} v & \omega \in B \\ t & \omega \in \overline{B} \end{cases} \quad \text{and} \quad X(\omega,f) = \begin{cases} v & \omega \in D \\ t & \omega \in \overline{D} \end{cases}$$

the preference $e \overset{A}{>} f$ is implied if $U(v) > U(t)$. However, we can not state a preference between a and e based on our development so far and therefore only a partial ordering can be determined. It seems intuitively clear, however, that if

$U(v) > U(r)$ and $U(t) > U(s)$ then $e \overset{A}{>} a$. This point will be considered later in section 5.0, so for the time being it is sufficient to consider the axiom as stated.

3.2 Implications of Axiom IV

Axiom IV, together with the previous axioms, requires the P-measure to have several properties. In this section we shall specify those which are used in subsequent sections.

Lemma 3.2.1. The P-measure is monotone on β, i.e., if $D \subset B$ then $P(D) \leqslant P(B)$.

Proof. By Assumption 3 there exists $a,b \in A$ such that for any reward $r,s \in R$ with $U(r) > U(s)$

$$X(\omega,a) = \begin{cases} r & \omega \in B \\ s & \omega \in \overline{B} \end{cases} \qquad X(\omega,b) = \begin{cases} r & \omega \in D \\ s & \omega \in \overline{D} \end{cases}.$$

By Axiom IIii, $a \overset{A}{\geqslant} b$, and this implies by Axiom IV that $P(B) \geqslant P(D)$.

Lemma 3.2.2. The P-measure is bounded by the inner and outer measure induced by μ, i.e., for any $D \in \beta$,

$$\mu_*(D) \leqslant P(D) \leqslant \mu^*(D)$$

where μ_* and μ^* are inner and outer measures respectively (see Appendix II for definition, p. 235).

Proof. Let $C \subset D \subset E$ be sets such that $D \varepsilon \beta$ and $C,E \varepsilon \Theta$. Since P-measure is monotone on β, $P(C) \leqslant P(D) \leqslant P(E)$. Thus, since $C,E \varepsilon \Theta$, we have $\mu(C) \leqslant P(D) \leqslant \mu(E)$, and the definition of outer and inner measures implies that

$$\mu_*(D) \leqslant P(D) \leqslant \mu^*(D).$$

Lemma 3.2.3. The P-measure agrees with the largest unique extension of μ such that the extension is a measure, i.e., $P(D) = \mu(D)$ if $D \varepsilon \overline{S}$ where $\overline{S} = \{D \subset \Omega | \mu_*(D) = \mu^*(D)\}$.

Proof. This follows directly from the previous lemma.

Lemma 3.2.4. The P-measure is "nearly additive"; that is, for any $C \varepsilon \beta$, $D \varepsilon \Theta$, such that $D \cap C = \emptyset$

1) $$|P(D \cup C) - P(D) - P(C)| \leqslant \mu^*(C) - \mu_*(C)$$

and for any $C \varepsilon \beta$ and $D \varepsilon \Theta$,

2) $$|P(C \cap D) + P(C \cap \overline{D}) - P(C)| \leqslant \mu^*(C) - \mu_*(C).$$

By "nearly additive" we mean, therefore, that given two disjoint sets, C and D such that $C \varepsilon \beta$, and $D \varepsilon \Theta$ then the sum $P(C) + P(D)$ can not differ from $P(D \cup C)$ by more than the difference of the outer and inner measure of C. Therefore, if the outer and inner measures are equal for the set C, then $P(D) + P(C) = P(D \cup C)$ for all sets $D \varepsilon \Theta$ such that $D \cap C = \emptyset$.

Proof. Two properties of inner and outer measures are (Appendix II, p. 235):
If $C \cap D = \emptyset$ and $C \varepsilon \beta$, $D \varepsilon \Theta$,

then $$\mu^*(D) + \mu^*(C) = \mu^*(D \cup C)$$

and $$\mu_*(D) + \mu_*(C) = \mu_*(D \cup C).$$

By lemma 3.2.2,

$$\mu_*(D \cup C) \leqslant P(D \cup C) \leqslant \mu^*(D \cup C),$$

or, using the above,

$$\mu_*(D) + \mu_*(C) \leqslant P(D \cup C) \leqslant \mu^*(D) + \mu^*(C).$$

Since $D \varepsilon \Theta$ which also implies that $\mu_*(D) = \mu^*(D) = P(D)$ subtraction yields

$$\mu_*(C) \leqslant P(D \cup C) - P(D) \leqslant \mu^*(C).$$

Combining this inequality with the following (from lemma 3.2.2,

$$\mu_*(C) \leqslant P(C) \leqslant \mu^*(C),$$

now yields

$$|P(D \cup C) - P(D) - P(C)| \leqslant \mu^*(C) - \mu_*(C).$$

Thus, P must be "nearly additive" for sets such that $D \cap C = \emptyset$ and $D \in \Theta$. Therefore P must be additive on $\bar{S}$.

Similarly, the inner and outer measures also satisfy the equalities

$$\mu_*(C) = \mu_*(C \cap D) + \mu_*(C \cap \bar{D})$$

and

$$\mu^*(C) = \mu^*(C \cap D) + \mu^*(C \cap \bar{D}).$$

By lemma 3.2.2 we have

$$\mu_*(C \cap D) \leqslant P(C \cap D) \leqslant \mu^*(C \cap D),$$

$$\mu_*(C \cap \bar{D}) \leqslant P(C \cap \bar{D}) \leqslant \mu^*(C \cap \bar{D}),$$

and hence

$$\mu_*(C \cap D) + \mu_*(C \cap \bar{D}) \leqslant P(C \cap D) + P(C \cap \bar{D}) \leqslant \mu^*(C \cap D) + \mu^*(C \cap \bar{D}).$$

Thus, with the equalities above, we have

$$\mu_*(C) \leqslant P(C \cap D) + P(C \cap \overline{D}) \leqslant \mu^*(C)$$

and, as before,

$$|P(C \cap D) + P(C \cap \overline{D}) - P(C)| \leqslant \mu^*(C) - \mu_*(C);$$

i.e., $P(C)$ must be "nearly" equal to $P(C \cap D) + P(C \cap \overline{D})$.

Savage's axioms imply Axiom IV. Of those approaches considered in Part I of the thesis, we recall that in the Savage approach a probability measure was derived on the subsets of β. The method used was substantially different, however, and it is not obvious that the P-measure specified here is consistent with that approach. Recall that Savage defines an ordering on all subsets of Ω (i.e., $\Theta = 2^{\Omega}$) by

$$B \overset{\Theta}{\geqslant} D \text{ if and only if } a \overset{A}{\geqslant} b$$

whenever $U(r) > U(s)$, $X_B(\cdot,a) = X_D(\cdot,b) = r$ and $X_{\overline{B}}(\cdot,a) = X_{\overline{D}}(\cdot,b) = s$.

Savage then proves that there exists a real valued function P on Θ such that

$$P(C) \geqslant P(D) \text{ if and only if } B \overset{\Theta}{\geqslant} D.$$

That is $P(C) \geqslant P(D)$ if and only if $a \overset{A}{\geqslant} b$, which implies

Axiom IV.

3.3 Comparison of different approaches using Ellsberg's paradox

Let us illustrate some properties of the evaluation function f which are stated in the first four axioms. It is appropriate to use the Ellsberg paradox II (Ellsberg, 1961) for this purpose, since empirical studies using this paradox have shown that most people differentiate between risk and uncertainty. (This point is discussed further in section 8.)

Suppose we are given an urn containing 30 red balls out of 90 balls, and the remaining 60 are an unknown mixture of yellow and black. We shall assume for simplicity that the probability of a ball of a given colour being drawn is equal to the proportional number of balls of that colour to the total number of balls.

Let A contain the following alternatives:

a) receiving \$100 if a red ball is drawn (event R)
receiving \$0 otherwise

b) receiving \$100 if a yellow ball is drawn (event Y)
receiving \$0 otherwise

c) receiving \$100 if a red or a black ball is drawn (event $B \cup R$)
receiving \$0 otherwise

d) receiving \$100 if a yellow or black ball is drawn (event $Y \cup B$)
receiving \$0 otherwise

e) receiving \$100 if a black ball is drawn (event B)
receiving \$0 otherwise

g) receiving \$0

Therefore by the statement of the problem we have the following:

$$A = \{a,b,c,d,e,g\} \qquad A_o = \{a,d,g\}$$

$$\Theta = \{\emptyset,\Omega,\{R\},\{B\cup Y\}\}$$

$$\beta = \{\emptyset,\Omega,\{R\},\{Y\},\{B\},\{R\cup Y\},\{R\cup B\},\{B\cup Y\}\}.$$

$$\mu(\emptyset) = 0$$

$$\mu(\Omega) = 1$$

$$\mu(R) = 1/3$$

$$\mu(B\cup Y) = 2/3.$$

Let $X(\cdot,i)$ denote the reward function corresponding to alternative i for any $i \in A$. By Axiom III the preference ordering on A_o (for any strictly increasing function U) must be $d \overset{A_o}{>} a \overset{A_o}{>} g$ since

$$EUX(\omega,d) > EUX(\omega,a) > EUX(\omega,g)$$

and $f(d) = EUX(\omega,d)$, $f(a) = EUX(\omega,a)$ and $f(g) = EU(\omega,g)$. All approaches suggested in the introduction agree with these values of f. Before we consider the properties f must have according to the axioms specified so far let us illustrate how the different approaches which were identified in the introduction extend the evaluation function f to A.

1) Savage approach

If we now consider the extension of μ on Θ to P on β such that P is a probability (as done by Savage, 1954, or Good, 1965) then P must be in the following form: $P(\emptyset) = 0$, $P(\Omega) = 1$, $P(R) = 1/3$, $P(Y) = p$, $P(B) = 2/3-p$. This would then imply that if a $\overset{A}{\geqslant}$ b then c $\overset{A}{\geqslant}$ d. In practice many decision makers would differentiate between risk and uncertainty and have the preference a $\overset{A}{>}$ b and d $\overset{A}{>}$ c. This approach therefore implies, a preference which is often contradicted by empirical studies.

2) Combination of evaluation function for uncertainty and risk

Consider the second method described in the introduction, i.e., a variation of the maximin.

Let Π denote the set of all possible combinations of black and yellow balls. Since the smallest possible number of yellow balls in the urn is zero, we have

$$\min_{\Pi} EUX(\omega, b) = U(O).$$

Since $U(X(\cdot, f)) = U(O)$, this would imply that we are indifferent between alternatives b and f, which all suggest must contradict most decision makers' preference.

3) Ellsberg approach

This method considers the most "likely" probability of a yellow ball being drawn. Ellsberg does not make it clear how this can be determined, but assumes for simplicity this means $P(Y) = P(B) = 1/3$.

The evaluation function would be (assuming $U(0) = 0$)

$f(a) = (1/3)U(100)$

$f(b) = (1-p_1)(1/3)U(100)$

$f(c) = (1-p_2)(1/3)U(100) + (1/3)U(100)$

$f(d) = (2/3)U(100)$

$f(e) = (1-p_3)(1/3)U(100)$

$f(g) = 0$

where p_i, $0 \leqslant p_i \leqslant 1$, $i=1,2,3$, is the "ambiguity" factor. Hence Ellsberg postulates that it is impossible to have a preference of $b \overset{A}{>} a$. In Ellsberg's approach, it would therefore be impossible to prefer uncertainty to risk. Although this is probably true in most cases, Ellsberg himself has suggested a decision problem where he predicts that uncertainty is preferred to risk[1].

---

[1]Consider: Urn I has 1,000 balls in it, each ball identified by a number, and each number from 1 to 1,000 represented. The probability of a random draw yielding a ball with a given number is 1/1,000. Urn II has 1,000 balls each identified by a number from 1 to 1,000. Thus, in the first urn all numbers from 1 to 1,000 are represented and in the second urn any number may appear zero, one or more times. The decision-maker is told that the number of occurrences of any given number is constrained only by the limits of 0 and 1,000 and by the fact that the total number of balls is 1,000. The decision-maker must decide which urn he prefers to draw from, given that he will win if any of n specified numbers is drawn and will lose nothing if a number not in the set of n is drawn. The subset of n numbers must come from the set of numbers from 1 to 1,000. Ellsberg suggests that, if n is very small, the decision maker will prefer Urn II where there may be as many as 1,000 balls

4) Fellner's approach

Fellner argues that actions in A_o can not be compared to actions in $A - A_o$. Hence we may compare b to e and a,d and g to each other, but b or d can not be compared. Fellner assumes the expected utility criterion can be used to order the alternatives in A_o, i.e., Fellner assumes Axiom III. For b,d an axiom similar to Axiom IV is used to determine a "P-measure" from which a probability measure is derived, and alternatives b and d are ordered by the expected utility of the derived probability. However, Fellner does not describe in his paper how alternatives such as c are handled where a mixture of known and unknown probabilities exist.

5) P-measure approach

In the theory developed here we have by Axiom III that if

$$a \overset{A}{>} b \text{ then } P(R) > P(Y) \text{ and if}$$

$$d \overset{A}{>} c \text{ then } P(B \cup Y) > P(B \cup R).$$

In the theory developed here we do not find this preference contradictory. However, this indicates that $P(B \cup Y)$ may not equal $P(B) + P(Y)$, or $P(B \cup R)$ may not equal $P(B) + P(Y)$ or

---

from the winning set, while Urn I has exactly n winning balls. Here ambiguity is considered to be favorable. As n increases, Ellsberg suggests that a point will be reached where the ambiguity associated with the action "draw from Urn II" is considered to be unfavorable.

(Becker and Brownson, 1964)

both. The set function P therefore may not be additive and hence not a measure.

This example shows therefore some of the difficulties of the methods proposed for the partial risk problem. It also shows that the P-measure is less restricted than the other approaches. There are, of course, some difficulties with the P-measure. For example, it may even be impossible to determine P(Y) if we do not assume a second urn (U_{II}) with a known proportion γ of red balls and with the appropriate extension of the σ-algebra Θ. The quantity P(Y) may be estimated by comparing alternative b with the following alternative:

h) receiving \$100 if a red ball is drawn from U_{II}
receiving \$0 otherwise

If b is preferred to h then $P(Y) > \gamma$ and similarly if h is preferred to b then $P(Y) < \gamma$ by Axiom III. By the transitivity assumption there exists a unique γ such that $\gamma = P(Y)$. To determine P(Y) by this method, we need the existence of U_{II} which we shall formalize as an assumption in section 3.3.

In Axiom III we assumed that $f(a) = \int UX(\cdot,a)d\mu$ for all $a \in A_o$. This definition uniquely defines the evaluation function $f(\cdot)$, primarily because μ is a measure. One can not directly extend the evaluation function $f(\cdot)$ to all actions in A, by simply replacing the measure μ by the P-measure P in the integral definition since it may not be uniquely defined. In order to specify a proper integral representation of $f(\cdot)$ additional assumptions regarding the P-measure must be made. This may present problems in that the desirable properties of P, e.g., non-additivity for some events, are to be retained.

For example, if we assume that $h \varepsilon A_O$, $\gamma < 1/3$ and $U(0) = 0$ then by replacing μ by P in the integral definition we obtain

$$f(h) = \gamma U(\$100) \quad \text{and}$$

$$f(e) = P(B)U(\$100).$$

Hence the decision maker acts as though his probability of obtaining \$100 is equal to P(B), and the probability of obtaining \$0 is equal to 1-P(B). So for the extension of the expected value using a P-measure is valid. However, if we extend the number of rewards to three in alternative k) we have

k) receiving \$100 if a black ball is drawn from U_I
   receiving \$100 + ε if a yellow ball is drawn from U_I
   receiving \$0 otherwise.

Assume for simplicity that P(Y) = P(B), and also that $a \overset{A}{>} b$.

By Axiom II, part ii, k > d. Calculating the "expected value" with P as a measure yields

$$U(100)\cdot 2/3 < U(100)P(B) + U(100 + \varepsilon)P(Y).$$

This implies

$$P(B) > \frac{2U(100)}{3(U(100 + U(100 + \varepsilon))}$$

Taking the limits as ε tends to zero, we obtain $P(B) \geqslant 1/3$, which would in turn imply $b \overset{A}{\geqslant} a$, contradicting our assumption.

It is obvious therefore that P can not be used to calculate the expected utility as if it were a measure.

3.4 Evaluation of P-measures

So far, we have only assumed the existence of the P-measure. There is no obvious method of determining what P(D) is equal to for an arbitrary $D \varepsilon \beta$. For example, if Θ only contains a finite number of sets, as in Ellsberg's paradox II, then if action a is preferred to action b, $0 < P(Y) < 1/3$. However P(Y) can not be specified completely since events for which the probability of occurrence is between 0 and 1/3 do not exist in this problem. Therefore to be able to specify the P-measure for any event we need to assume the existence of Urn II or that if α is a real number between zero and one, then there exists an event $C \varepsilon \Theta$ such that $\mu(C) = \alpha$, and also that if a reward function is defined by

$$X(\omega,a) = \begin{cases} r & \omega \varepsilon C \\ s & \omega \varepsilon \overline{C} \end{cases}$$

then $a \varepsilon A_o$.

Therefore, for any $D \varepsilon \beta - \Theta$, the P-measures of D can easily be determined by comparing actions a to b where

$$X(\omega,b) = \begin{cases} r & \omega \varepsilon D \\ s & \omega \varepsilon \overline{D} \end{cases}$$

Thus, by varying α until we become indifferent between a and b, we will determine a specific α such that $P(D) = \alpha$.

A second difficulty arises in the development of the theory because the reward set R is defined as being an arbitrary set; that is, we have not specified its cardinality. It may be finite, countable or even uncountable. We shall assume that the reward set R is not finite; that is, it may be countable or uncountable. In that way we do not restrict the action set A in any way. These assumptions do not indicate any ordering preference among the actions in $A - A_o$, that is, it only specifies some conditions on decision problems under risk and therefore ought not to influence our preference under partial risk problems.

Assumption 4. i) If $Y(\cdot)$ is any function from Ω to R such that $UY(\cdot)$ is Borel measurable with respect to Θ, then there exists $a \in A$ such that $X(\cdot,a) = Y(\cdot)$.

ii) There exists an action $a \in A_o$ such that

$$\mu(UX(\omega,a) < \gamma) = \frac{\gamma - U(r_o)}{U(r^o) - U(r_o)}$$

for each $\gamma \in [U(r_o), U(r^o)]$.

Part i) of the assumption implies that if Y is a function from Ω to R for which EUY is defined then there exists $a \in A_o$ such that $Y(\cdot) = X(\cdot,a)$. Part ii) is more general than the assumption discussed previously, that is, for every α between zero and one there exists an event $C \in \Theta$ such that $\mu(C) = \alpha$. The assumption that there exists an action a which has a uniform distribution on $[U(r_o), U(r^o)]$ implies that if α is any number between (0,1) and $C = \{\omega : UX(\omega,a) < \gamma\}$ where

$$\gamma = \alpha\left[U(r^{o}) - U(r_{o})\right] + U(r_{o})$$

then $P(C) = \alpha$. However, the converse does not necessarily hold. That is, if we assume for any $\alpha \varepsilon (0,1)$ there exists a $C \varepsilon \Theta$ such that $\mu(C) = \alpha$, then this does not imply that there exists a reward function with uniform distribution. Part ii) of the assumption also implies that the rewards are at least countable, not finite.

4.0 Sequences of reward functions

In section 3 we showed that the P-measure is monotone which is, of course, also one of the properties of a measure. In this section we shall make the additional assumption that P-measure is continuous from below, that is, if $B_1, B_2, \ldots$ is an increasing sequence of sets in β, then

$$\lim P(B_i) = P(\lim B_i).$$

This holds as a property of a measure also. (See, for example, Halmos, 1950, Theorem E, pp 38.)

Actually, we shall make the assumption slightly stronger in terms of reward functions rather than in terms of the measure. For most practical purposes this axiom is not needed since it is usually sufficient to consider only a finite number of states of nature, or at most finitely many sets in β. It is needed, however, for a consistent mathematical development. In section 4.1 we shall state the axiom, and in section 4.2 we shall consider some of its implications.

4.1 Statement of the axiom

Before stating the axiom, we shall need an additional definition of what is meant by "convergence of a sequence of reward functions" or "convergence of a sequence of actions". The normal way would be to define topologies on A and R, and describe convergence in terms of these topologies. For our purposes it is sufficient to use the following:

Definition. If a_n is a sequence of actions in A, then we shall say a_n converges to a (or a sequence of reward functions $X(\cdot,a_n)$ converges to a reward function $X(\cdot,a)$) if

$$\lim_n U(X(\omega,a_n)) = U(X(\omega,a)) \qquad \text{for all } \omega \in \Omega.$$

In mathematical terms we have induced a topology on A by the function $U(X(\cdot,a))$ from the natural topology on the real line. If a_n converges to a, we shall denote this by $\lim a_n = a$ or $a_n \to a$. Similarly the convergence of $X(\cdot,a_n)$ to $X(\cdot,a)$ is denoted by

$$\lim_n X(\cdot,a_n) = X(\cdot,a) \text{ or } X(\cdot,a_n) \to X(\cdot,a).$$

Let $X(\cdot,a_n)$ be a sequence of reward functions, such that $UX(\cdot,a_n)$ converges to a function $Y(\cdot)$. If $a_n \varepsilon A_o$ for all n then $Y(\cdot)$ is a Θ-measurable function. However, there may not exist an $a \varepsilon A_o$ such that $Y(\cdot) = UX(\cdot,a)$. For example if $B \varepsilon \Theta$, R is the open interval (0,1) and $U(x) = x$ then the sequence of functions defined by

$$X(\omega,a_n) = \begin{cases} 1-1/n & \omega \varepsilon B \\ 1/n & \omega \varepsilon \overline{B} \end{cases}$$

converges to the function

$$Y(\omega) = \begin{cases} 1 & \omega \varepsilon B \\ 0 & \omega \varepsilon \overline{B}. \end{cases}$$

It is clear that there does not exist an $a \in A_o$ such that $Y(\cdot) = X(\cdot,a)$. This will lead to some difficulties in section 5. We shall therefore make the following assumption:

<u>Assumption 5</u>. If a_n is a sequence in A_o such that $a_n \to a$ then $a \in A_o$.

Assumption 5 therefore states that in a certain topology that A_o is a closed set. Equivalent assumption would be to assume R to be a closed set, and state the assumption in terms of a reward function.

If a sequence of actions a_n therefore converges to a, the reward $X(\omega,a_n)$ becomes closer to the reward $X(\omega,a)$ for all $\omega \in \Omega$, as n tends to infinity and the two reward functions become indistinguishable. It seems intuitively obvious then that $f(a_n)$ should become closer to $f(a)$ as n increases. Axiom V formalizes this intuition.

<u>Axiom V</u>. If $X(\cdot,a_n)$ n=1,2,... is a sequence of reward functions such that $b \overset{A}{\geqslant} a_{n+1} \overset{A}{\geqslant} a_n$ for n=1,2,... and $\lim X(\omega,a_n) = X(\omega,b)$ for all $\omega \in \Omega$, then for any action c for which $b \overset{A}{>} c$ there exists an N such that $a_n \overset{A}{\geqslant} c$ for all $n > N$.

As an example of the implication of Axiom V consider the following example: Let α be an arbitrary number from the interval [0,1]. We shall say that the event B_n occurred if the number chosen belongs to the interval [0,3/4-1/n) where $n > 2$. Similarly we say the event B occurred if the number chosen belongs to the interval [0,3/4). Consider the following alternatives:

a_n) receiving \$100 if B_n occurs
receiving \$0 otherwise

b) receiving \$100 if B occurs
receiving \$0 otherwise

As n increases, the set B_n becomes closer to the set B and, therefore

$$\lim X(\cdot,a_n) = X(\cdot,b).$$

Therefore if n is very large, Axiom V assumes that we would be "nearly" indifferent between a_n and b. That is, let an additional alternative c be defined by:

c) receiving \$100 if C occurs
receiving \$0 otherwise.

If $b \overset{A}{>} c$ in the strict sense (that is, $P(B) > P(C)$) and if a_n is "nearly" indifferent to b, Axiom V asserts that a_n is also preferred to c, i.e., $a_n \overset{A}{\geqslant} c$.

4.2 Implications of the axiom

The implications of this axiom together with previous assumptions are very strong from a mathematical viewpoint. We shall prove two of these implications here. The first implies that the P-measure is continuous, that is, $\lim P(B_i) = P(\lim B_i)$ for any increasing sequence of sets in β. The second implication is that if $a_n \to a$ where $a \overset{A}{\geqslant} a_{n+1} \overset{A}{\geqslant} a_n$ for all n, then $f(a_n) \to f(a)$. This implies therefore that if $a_n \in A_o$ and $a \in A_o$, that

$$\int UX(\omega,a)\,d\mu = \int \lim UX(\omega,a_n)\,d\mu = \lim \int UX(\omega,a_n)\,d\mu .$$

We shall also show that in Savage's approach this axiom is not satisfied.

Several implications are now verified.

<u>Lemma 4.2.1</u>. If B_n is a sequence of increasing sets in β such that $\lim B_n = B$, then $\lim P(B_n) = P(B)$.

<u>Proof</u>. For any $r,s \in R$ such that $U(r) > U(s)$, consider the reward functions

$$X(\omega,a_n) = \begin{cases} r & \omega \in B_n \\ s & \omega \in \overline{B}_n \end{cases} \quad \text{and} \quad X(\omega,b) = \begin{cases} r & \omega \in B \\ s & \omega \in \overline{B} \end{cases}$$

for each $n=1,2,\ldots$. Then $b \overset{A}{\geqslant} \ldots \overset{A}{\geqslant} a_3 \overset{A}{\geqslant} a_2 \overset{A}{\geqslant} a_1$ by Axiom II, and by Axiom IV(b) this implies that

$$P(B_1) \leqslant P(B_2) \leqslant P(B_3) \ldots \leqslant P(B).$$

That is, $P(B_n)$ is an increasing sequence with an upper bound. This implies that $\lim P(B_i) = \alpha$ exists. If $\alpha < P(B)$ then by Assumption 4, there exists a $C \in \Theta$ such that $\alpha < P(C) < P(B)$, and also there exists an action $c \in A_o$ such that

$$X(\omega,c) = \begin{cases} r & \omega \in C \\ s & \omega \in \overline{C} \end{cases}.$$

Since $b \overset{A}{>} c$ and $c \overset{A}{>} a_n$ for all n by Axiom IV and $a_n \to b$ by definition which contradicts Axiom V. Hence $\lim P(B_i) = P(B)$.

Lemma 4.2.2. If $X(\cdot, a_n)$ $n=1,2,\ldots$ is a sequence of reward functions such that $a \overset{A}{\geqslant} a_{n+1} \overset{A}{\geqslant} a_n$ for $n=1,2,\ldots$ and $a_n \to a$ $X(\Lambda,\omega)$ then $f(a_n) \to f(a)$.

Proof. Assume $\lim f(a_n) = \alpha < f(a)$ then by Assumption 3 there exists $c \in A_o$ defined by the reward function

$$X(\omega,c) = \begin{cases} r^o & \omega \in B \\ r_o & \omega \in \overline{B} \end{cases}$$

where

$$\mu(B) = \frac{\frac{f(a)+\alpha}{2} - U(r_o)}{U(r^o)-U(r_o)} .$$

Hence

$$f(c) = U(r_o) + [U(r^o) - U(r_o)] \frac{\frac{f(a)+\alpha}{2} - U(r_o)}{U(r^o)-U(r_o)}$$

$$= \frac{f(a)+\alpha}{2} .$$

Therefore, $a \overset{A}{>} c \overset{A}{>} \lim a_n$ contradicting Axiom V.

Savage approach does not satisfy Axiom V. Although Savage did not explicitly argue against this axiom, his axioms contradict it. To see this, let us assume that we start with a given measure space (Ω,Θ,μ) where μ is an additive measure. Using the first six axioms in his approach, Savage proved that there exists a function U from R to the real line such that for any simple function (see Appendix II for definition)

$$f(a) = EUX(\cdot,a) = \sum_{i=1}^{n} UX_{B_i}(\cdot,a)\mu(B_i)$$

where $UX_{B_i}(\cdot,a)$ is a constant for each $i=1,\ldots n$.

Since Savage defines Θ as all subsets of Ω, there does not exist, in general, a σ-additive set function μ on Θ. Savage does not restrict Ω or Θ in his approach, but rather assumes that μ is only finitely additive. We shall show that if Axiom V were accepted then the measure μ must be σ-additive and hence the general assumption of Ω and Θ can not be made.

If B_i $i=1,\ldots$ is a sequence of disjoint sets, we are required to have

$$\sum_{i=1}^{\infty} \mu(B_i) = \mu(B) \text{ where } UB_i = B.$$

Since in Savage's approach there always exist rewards such that $U(r) = 1$ and $U(s) = 0$, the following reward functions may be considered:

$$UX(\omega,a) = \begin{cases} 1 & \omega \varepsilon B \\ 0 & \omega \varepsilon \overline{B} \end{cases} \quad \text{and} \quad UX(\omega,a_n) = \begin{cases} 1 & \omega \varepsilon \sum_{i=1}^{n} B_i \\ 0 & \text{otherwise} \end{cases}$$

Then, $a \overset{A}{\geqslant} a_n$ for all n by Axiom II, and

$$\lim X(\cdot,a_n) = X(\cdot,a).$$

It is also clear that $X(\cdot,a)$ and $X(\cdot,a_n)$ $n=1,2,\ldots$ belong to A_o.

The expected value of $X(\cdot,a)$ is equal to $\mu(B)$, and the expected value of $X(\cdot,a_n)$ is equal to $\mu(\sum_{i=1}^{n} B_i) = \sum_{i=1}^{n}\mu(B_i)$, by finite additivity.

If Axiom V holds,

$$\mu(B) = \lim\mu(\sum_{i=1}^{n} \mu(B_i)) = \sum_{i=1}^{\infty} \mu(B_i),$$

and hence μ is σ-additive.

Therefore if we assume that a sequence of reward functions satisfies Axiom V, we must also have σ-additivity of the measure if finite additivity is assumed. This implies, therefore, that the set of the states of nature, i.e., the set Ω, must be sufficiently small so that all subsets may be made measurable if Savage accepts Axiom V. Since Savage does not restrict Ω, or A for that matter, and assumes finite additivity, we can only conclude that Savage rejects Axiom V.

5.0 Stochastic dominance axiom

In previous sections of Part II we have discussed the properties of the evaluation function $f(\cdot)$ for actions which result in only two rewards. So far the axioms give only a partial ordering on the set of actions. In this section we shall specify one additional axiom which will allow $f(\cdot)$ to be specified completely. In section 5.1 we shall state the axiom and in section 5.2 we shall consider some of its implications.

5.1 Statement of the axiom

Axiom VI can be considered as a generalization of Axiom IVb and will, in fact, replace it. Recall that Axiom IVb states that if the actions $a,b \in A$ are defined by the reward functions

$$X(\omega,a) = \begin{cases} r & \omega \in B \\ s & \omega \in \overline{B} \end{cases} \qquad X(\omega,b) = \begin{cases} r & \omega \in D \\ s & \omega \in \overline{D} \end{cases}$$

where $U(r) > U(s)$, then

$$a \overset{A}{\geqslant} b \quad \text{if and only if} \quad P(B) \geqslant P(D).$$

If we now consider the case where $a,b \in A$ are defined by

$$X(\omega,a) = \begin{cases} r & B_1 \\ s & B_2 \\ t & B_3 \end{cases} \qquad X(\omega,b) = \begin{cases} r & D_1 \\ s & D_2 \\ t & D_3 \end{cases}$$

for $B_i \cap B_j = \emptyset$, $D_i \cap D_j = \emptyset$ $i \neq j$, then in some cases the previous axioms would be sufficient to specify the preference between a and b. For example, if $D_1 \subset B_1$, $D_1 \cup D_2 \subset B_1 \cup B_2$ and $U(r) > U(s) > U(t)$, then the dominance axiom (Axiom II) would specify the preference $a \overset{A}{\geqslant} b$. Axiom IVc generalized Axiom II by not requiring that $D \subset B$, but only that $P(B) \geqslant P(D)$. In the same way we shall generalize Axiom VI, using Axiom IIb which, in this case, requires that $a \overset{A}{\geqslant} b$ if and only if $P(B_1) \geqslant P(D_1)$ and $P(B_1 \cup B_2) \geqslant P(D_1 \cup D_2)$. Analogously, in terms of two arbitrary actions $a,b \in A$, this would require that $a \overset{A}{\geqslant} b$ if and only if $P(X(\omega,a) > r) \geqslant P(X(\omega,b) > r)$ for all $r \in R$.

However, it is more convenient to consider $P(UX(\omega,a) > \alpha)$ rather than $P(X(\omega,a) > r)$, since the function $F(\alpha,a) \equiv 1-P(UX(\omega,a) > \alpha)$ defines a probability distribution on the real line, as we shall show in lemma 5.2.1. Hence, the desired property can be restated more simply as $a \overset{A}{\geqslant} b$ if and only if $F(\alpha,a) \leqslant F(\alpha,b)$ for all real numbers α. This concept is not new in the theory of expected utility theory. However in most cases R is considered to be the real line, and in this case the distribution function $F(\alpha,a)$ of the reward function $X(\cdot,a)$ is defined in the normal way, that is

$$F(\alpha,a) = P(X(\cdot,a) \leqslant \alpha).$$

In this case the following theorem can be stated:

<u>Theorem</u>. (L. Tesfatsion, 1974). A necessary and sufficient condition for $EUX(\cdot,a) \geqslant EUX(\cdot,b)$, where U is any nondecreasing bounded utility function is

that $F(\alpha,a) < F(\alpha,b)$ for every real number α.

In terms of the ordering on A_o this states therefore that $a \overset{A_o}{\geqslant} b$ if $F(\alpha,a) \leqslant F(\alpha,b)$. Axiom VI extends this property to all actions in A.

Axiom VI. If $P(UX(\omega,a) > \alpha) \geqslant P(UX(\omega,b) > \alpha)$ for all real numbers α then $a \overset{A}{\geqslant} b$ and if $P(UX(\omega,a) > \alpha) > P(UX(\omega,b) > \alpha)$ for some α then $a \overset{A}{>} b$.

5.2 Implications of Axion VI

In this section we shall prove two lemmas which give the foundation for specifying $f(\cdot)$ completely. The most important implication of the assumptions so far is that the P-measure induces a distribution function on the real line.

Lemma 5.2.1. If $F(\alpha,a) = 1-P(\omega:UX(\omega,a) > \alpha)$ then $F(\alpha,a)$ is a distribution function.

Proof. We must show that $F(\cdot,a)$ satisfies

i) Non-negativity, i.e., $F(\cdot,a) \geqslant 0$.

ii) Continuity from above, i.e., if α_i is a decreasing sequence of real numbers converging to α_o then $F(\alpha_i,a)$ converges to $F(\alpha_o,a)$.

iii) If α_j is a decreasing sequence of real numbers, $\alpha_j \to -\infty$ then $F(\alpha_j,a) \to 0$.

iv) If α_j is an increasing sequence of real numbers $\alpha_j \to +\infty$ then $F(\alpha_j,a) \to 1$.

v) If $\alpha_1 \leqslant \alpha_2$, then $F(\alpha_1,a) \leqslant F(\alpha_2,a)$.

Since the P-measure is always bounded between 0 and 1 (by Axiom IVa and Axiom II) this implies that $F(\alpha,a)$ is always greater than or equal to zero. To show ii), let α_i be a decreasing sequence of real numbers converging to α_o. Define $B_n = (\omega:U(X(\omega,a)) > \alpha_n)$ then B_n is an increasing sequence of sets in β, such that $\lim B_n = B = \{\omega:U(X(\omega,a) > \alpha_o\}$. By lemma 4.2.1 section 4.2 (i.e., by Axiom V) therefore

$$\lim P(B_n) = P(B)$$

or $$\lim F(\alpha_n,a) = 1-\lim P(\omega:U(X(\omega,a)) > \alpha_n)$$

$$= 1-P(\omega:U(X(\omega,a)) > \alpha_o)$$

$$= F(\alpha_o,a).$$

Properties (iii) and (iv) now follow directly since $U(r_o) \leqslant U(X(\omega,a)) \leqslant U(r^o)$; that is, if $\alpha < U(r_o)$ then $P(\omega:U(X(\omega,a)) > \alpha) = 1$ and if $\alpha > U(r^o)$, $P(\omega:U(X(\omega,a)) > \alpha)=0$, (v) follows directly from lemma 3.2.1.

A standard result in probability theory is that $\int \alpha dF(\alpha,a) = EUX(\cdot,a)$ (see, for example, Theorem 1.6.12 in Ash (1972)). This implies, therefore, that the ordering on A_o must satisfy the numerical ordering of $\int \alpha dF(\alpha,a)$ for all $a \in A_o$. We shall state this result as a lemma.

Lemma 5.2.2. Let $a,b \in A_o$. If $a \overset{A_o}{\geqslant} b$, then $\int \alpha dF(\alpha,a) \geqslant \int \alpha dF(\alpha,b)$.

Our objective now is to show that the ordering on A must also satisfy $a \overset{A}{\geqslant} b$ implies $\int \alpha dF(\alpha,a) \geqslant \int \alpha dF(\alpha,b)$.

We shall do this in two steps.

First, we shall show that if $F(\alpha,b)$ is any distribution with $b \in A$, there exists an $a \in A_o$ such that $F(\alpha,b) = F(\alpha,a)$. This is stated in Theorem 5.2.3.

Theorem 5.2.3. Let $F(\alpha,a)$ be the distribution function induced by $X(\cdot,a)$ for any $a \in A-A_o$. Then there exists a $b \in A_o$ such that $F(\alpha,a) = F(\alpha,b)$ for all real numbers α.

Proof. To prove this we shall construct a sequence of measurable functions which converge to a reward function with the required distribution function.

Let $S_n = \{\gamma_o,\ldots,\gamma_m\}$ be a strictly increasing set of numbers in the interval $[U(r_o),U(r^o)]$ such that $\gamma_o = U(r_o)$, $\gamma_m = U(r^o)$ and

$$F(\gamma^-_{i+1},a) - F(\gamma_i,a) \leq \frac{1}{n} \quad \text{for } i=1,\ldots,m-1$$

where $F(\gamma^-,a) = \lim\limits_{\substack{\alpha<\gamma \\ \alpha\to\gamma}} F(\alpha)$.

Since F is right-continuous such a sequence must always exist.

Let T_n denote the set of all discontinuity points of $F(\cdot,a)$ such that if $y \in T_n$ then

$$F(y,a) - F(y^-,a) > \frac{1}{n}.$$

It is obvious then that $T_n \subset S_n$.

The next step in the proof is to define a measurable reward function with a distribution function which approximates $F(\alpha,a)$. To do so, we must refer back to Assumption 4. This assumption specifies the existence of a reward function $UX(\cdot,c)$ with a uniform distribution. Define the sets

$$B_o = \{\omega : UX(\omega,c) \leqslant F(\gamma_1^-,a)\left[U(r^o) - U(r_o)\right] + U(r_o)\}.$$

$$B_i = \{\omega : F(\gamma_i,a)\ U(r^o) - U(r_o)\ + U(r_o) < UX(\omega,c) \leqslant$$

$$< F(\gamma^-_{i+1},a)\left[U(r^o) - U(r_o)\right] + U(r_o)\}$$

$$\text{for all } \gamma_i \in S_n - T_n, \quad i=1,2,\ldots,m-1.$$

$$C_i = \{\omega : F(\gamma_i,a)\left[U(r^o) - U(r_o)\right] + U(r_o) < UX(\omega,c) \leqslant$$

$$\leqslant F(\gamma^-_i,a)\left[U(r^o) - U(r_o)\right] + U(r_o)\}$$

$$\text{for all } \gamma_i \in T_n, \quad i=0,1,2,\ldots,m.$$

Then the set $B_o, B_1, \ldots$ and $C_1, \ldots$ are all pairwise disjoint, and if $\gamma_i \in S_n$ and $i \geqslant 1$ then

$$\mu(B_i) = F(\gamma^-_i,a) - F(\gamma_{i-1})a$$

and also if $\gamma_i \in T_n$ then

$$\mu(C_i) = F(\gamma_i) - F(\gamma^-_i).$$

Assumption 3 also implies that there exists $\omega_i \in \Omega$ such that

$$UX(\omega_i,c) \in [\gamma_i,\gamma_i+1/n) \cap [\gamma_i,\gamma_{i+1}] \quad \text{for } i=0,\ldots,m-1$$

and also $\omega_m \in \Omega$ such that

$$UX(\omega_m,c) \in (\gamma_m-1/n,\gamma_m] \cap (\gamma_{m-1},\gamma_m].$$

Define a reward function

$$UX(\omega,b_n) = \begin{cases} UX(\omega_i,\alpha) & \omega \in B_i \quad i=0,1,\ldots,m-1 \\ \gamma_i & \omega \in C_i \quad i=1,2,\ldots,m \end{cases}$$

Then $b_n \in A_o$ by Assumption 4i since $UX(\cdot,b_n)$ is a measurable function. Then for any $\alpha \in [U(r_o),U(r^o)]$ we have:

<u>Case 1</u>. $\alpha \in T_n$

$$P(UX(\cdot,b_n) < \alpha) = F(\alpha,a).$$

<u>Case 2</u>. $\alpha \notin T_n$, then there exists an interval $[\gamma_{i+1},\gamma_i)$ or $(\gamma_{m-1},\gamma_m]$ to which α belongs, say $[\gamma_{i+1},\gamma_i)$. Then

$$F(\gamma_i,a) \leqslant P(UX(\cdot,b_n) \leqslant \alpha) \leqslant F(\gamma_{i+1},a)$$

and therefore

$$|P(UX(\cdot,b_n) \leqslant \alpha) - F(\alpha,a)| \leqslant F(\gamma_{i+1},a) - F(\gamma_i,a) \leqslant \frac{1}{n}.$$

Since γ_n is strictly increasing $UX(\omega,b_n)$ converges as n increases and by Assumption 4 there exists $b \varepsilon A_o$ such that b_n converges to b. This completes the proof.

This theorem therefore completes the basic idea behind Part II of the thesis. Our arguments are as follows: Suppose we are interested in comparing actions a_1 and b_1 where $a_1, b_1 \varepsilon A-A_o$. By theorem 5.2.3 there exists an $a,b \varepsilon A_o$ such that $F(\alpha,a_1)$ and $F(\alpha,b_1)$ have the same distribution as $F(\alpha,a)$ and $F(\alpha,b)$. Axiom VI then specifies that $a_1 \overset{A}{=} a$ and $b_1 \overset{A}{=} b$. Therefore the preference between a_1 and b_1 can be determined by the preference between a and b, which can easily be done since these satisfy the expected utility criterion. In section 6 we shall formalize this argument.

6.0 Summary of assumptions and axioms and the basic results

In this section we shall summarize the assumptions and the axioms so far made and we shall also show in theorem 6.0.1 that they are sufficient to specify the evaluation function $f(\cdot)$ for all $a \varepsilon A$. The only difference between the axioms here and in previous sections is in Axiom IV, part b, which has been replaced by Axiom VI.

The following five assumptions were made:

Assumption 1. For each $r \varepsilon R$, there exists an $a \varepsilon A_o$ such that $X(\omega,a) = r$ for all $\omega \varepsilon \Omega$.

Assumption 2. i) There exist r_o and r^o in R, such that for each $r \varepsilon R$

$$r^o \overset{R}{\geqslant} r \overset{R}{\geqslant} r_o$$

ii) The function $UX(\cdot,a)$, $a \varepsilon A_o$ is a Borel measurable function with respect to Θ. The function $UX(\cdot,a)$, $a \varepsilon A$ is a Borel measurable function with respect to β.

Assumption 3. For any $D \varepsilon \beta$ and for any $r,s \varepsilon R$ there exists $b \varepsilon A$ such that

$$X(\omega,b) = \begin{cases} r & \omega \varepsilon D \\ s & \omega \varepsilon \overline{D} \end{cases}.$$

Assumption 4. i) If $Y(\cdot)$ is any function from Ω to R such that $UY(\cdot)$ is Borel measurable with respect to Θ, then there exists $a \varepsilon A$ such that $X(\cdot,a) = Y(\cdot)$.

ii) There exists an action $a \in A_o$ such that

$$\mu(UX(\omega,a) < \gamma) = \frac{\gamma - U(r_o)}{U(r^o) - U(r_o)}$$

for each $\gamma \in [U(r_o), U(r^o)]$.

<u>Assumption 5</u>. If a_n is a sequence in A_o such that $a_n \to a$ then $a \in A_o$.

The following six axioms were made:

<u>Axiom I</u>. There exists a real-valued function $f(\cdot)$ on A such that

$$a \overset{A}{\geqslant} b \text{ if and only if } f(a) \geqslant f(b).$$

<u>Axiom II</u>. i) If for any $r \in R$ there exists $a,b \in A$ such that $X(\omega,a) = r$ and $X(\omega,b) = r$ for all $\omega \in \Omega$ then $f(a) = f(b)$.

ii) If $X(\cdot,a)$ dominates $X(\cdot,b)$ then $a \overset{A}{>} b$.

<u>Axiom III</u>. If $b \in A_o$ then $f(b) = EUX(\cdot,b)$.

<u>Axiom IV</u>. There exists a set function P defined on β such that if $B \in \Theta$ then $P(B) = \mu(B)$.

<u>Axiom V</u>. If $X(\cdot,a_n)$ n=1,2,... is a sequence of reward functions such that $b \overset{A}{\geqslant} a_{n+1} \geqslant a_n$ for n=1,2,... and

$\lim X(\omega,a_n) = X(\omega,b)$ for all $\omega \in \Omega$ then for any actions c for which $c \overset{A}{<} b$ there exists an N such that $c \overset{A}{<} a_n$ for all $n > N$.

<u>Axiom VI</u>. If $P(UX(\omega,a) > \alpha) \geqslant P(UX(\omega,b) > \alpha)$ for all real numbers α then $a \overset{A}{\geqslant} b$, if in addition $P(UX(\cdot,a) > \alpha) > P(UX(\cdot,b) > \alpha)$ for some α then $\alpha \overset{A}{>} b$.

Theorem 6.0.1 extends lemma 5.2.2 in which we showed that

$$f(a) = \int \alpha dF(\alpha,a) \quad \text{for all } a \in A_o$$

where $F(x,a) = 1 - P\{\omega : UX(\omega,a) > x\}$. Here it will be shown that this is the case for all $a \in A$.

<u>Theorem 6.0.1</u>. If the Axioms I-VI and the Assumptions 1-5 stated above hold, then $f(a) = \int \alpha dF(\alpha,a)$ for all $a \in A$.

<u>Proof</u>. It is sufficient to show that this holds for $a \in A - A_o$. If $a \in A - A_o$ then by Theorem 5.2.3 there exists $b \in A_o$ such that $F(\alpha,a) = F(\alpha,b)$ for all real numbers α. This implies that

$$P(\omega : U(X(\omega,a) > \alpha) \geqslant P(\omega : U(X(\omega,b)) > \alpha)$$

for all real numbers α, so by Axiom VI $a \overset{A}{\geqslant} b$. Similarly,

$$P(\omega : U(X(\omega,b) > \alpha) \geqslant P(\omega : U(X(\omega,a)) > \alpha),$$

for all real numbers α, hence $b \overset{A}{\geqslant} a$. Therefore $a \overset{A}{=} b$

or $f(a) = f(b)$. Since $f(b) = \int x dF(x,b) = \int x dF(\cdot,a) =$ $f(a)$ the proof is completed.

To summarize the results, we have shown that for a given utility function U, and for any actions $a,b \in A$ such that

$$a \overset{A}{\geqslant} b \text{ then } \int \alpha dF(\alpha,a) \geqslant \int \alpha dF(\alpha,b).$$

One of the properties of expected utility is that any affine transformation of the utility function would satisfy the same ordering on A_o, that is, it is immaterial if we use $U(X)$ or $\alpha U(X)+\gamma$ for any $\alpha > 0$. In the development here we have made use of a specific utility function and for the theory to be reasonably useful we must show that for any affine utility function the same ordering will be obtained.

Therefore we must show that $\int zdF(z,a) > \int zdF(z,b)$ where $F(z,i) = 1-P\{\omega{:}U(X(\cdot,i)) > z$ $i=a,b$ implies that $\int zdF'(z,a) > \int zdF'(z,b)$ where $F'(z,i) = 1-P(\omega{:}\alpha U(X(\cdot,i)) + \gamma > z)$ for $\alpha > 0$ and γ any real number. Since $F'(z,i) =$ $F((z-\gamma)/\alpha,i)$ $i=a,b$ the following identity holds:

$$\int zdF' = \alpha \int zdF + \gamma.$$

We have that if $\int zdF(z,a) > \int zdF(z,b)$, then

$$\alpha \int zdF(z,a) + \gamma > \alpha \int zdF(a,b) + \gamma$$

or

$$\int zdF'(z,a) > \int zdF'(z,a)$$

for any $\alpha > 0$ and any γ.

Therefore the same transformation of U would preserve the ordering on A.

6.1 Relation of the P-measure approach to alternative approaches

In the introduction we specified several approaches to the partial risk problem. We shall consider some of these in relation to the approach developed here. The P-measure as a probability will be considered in section 8.0.

In two of the approaches, a set Π of possible probability measures on β must be specified. Hence existence of the probability measure is taken for granted. This is not an obvious assumption for a theoretical development, but for practical problems this is not a serious drawback.

One of the criteria is then specified by choosing the alternative $a \in A$ for which $\min_{\Pi} EU(X(\cdot,a))$ is maximum. This approach therefore ignores the possibility that one of the distributions in Π may be more likely than another. For example, in Ellsberg's paradox, we may be told that there is more than a .50 chance that there are more black than yellow balls. This type of information can not be used in this approach.

In order to relate this to Ellsberg's paradox II again, consider the following alternatives:

a) receiving \$100 if a black ball is drawn
receiving \$0 otherwise

Then, since there may be zero black balls

$$\min_{\Pi} EU(X(\cdot,a)) = U(0).$$

This is equivalent to setting the P-measure equal to μ_*\{black ball is drawn\}, the inner measure of the event in question.

In the more general case, if $U(r_1) > U(r_2) > \ldots > U(r_n)$ and alternative b is defined by

b) receiving r_1 if B_1 occurs
receiving r_2 if B_2 occurs
. . .
receiving r_n if B_n occurs

where $B_i \cap B_j = \emptyset$, $i \neq j$.

Then the P-measures would be defined as

$$\begin{aligned} P(B_1) &= \mu_*(B_1) \\ P(B_2) &= \mu_*(B_2) \\ &\vdots \\ P(B_n) &= 1 - \sum_{i=1}^{n-1} P(B_i). \end{aligned}$$

Hence the approach which uses $\min_{\Pi} EU(X(\cdot,b))$ as a decision criterion is a special case of the P-measure. However, it seems unlikely that anyone would have such an extreme P-measure.

In Ellsberg's approach we must determine the "most likely" linear combinations of distributions in Π. This combination is denoted by $\hat{P}$ and it is not obvious how it can be determined. Let us assume that this can be determined and then the evaluation function is specified by

$$f(a) = \rho\int UX(\cdot,a)d\hat{P} + (1-\rho)\min_{\Pi} EUX(\cdot,a)$$

where ρ is an "ambiguity factor". It is obvious that ρ must be a function of the "ambiguity" of receiving certain rewards, rather than a function of ambiguity of the probabilities of certain events.

The "ambiguity factor" thus becomes a function on the actions, i.e., each action may have a different "ambiguity factor" specified by the decision maker. Therefore this method is no different from the Hurwicz α-criterion, where α is the "ambiguity factor", which simply defines for each action

$$f(a) = \alpha U(r_o) + (1-\alpha)U(r^o)$$

where α depends on the action. Although this evaluation function can certainly be used to specify an ordering on A given Axiom I and Assumption 1, it gives little understanding of how to make decisions.

In order for Ellsberg's criterion to be useful, some method has to be found to determine the "ambiguity factor" ρ, independently of the action. Becker and Brownson (1964) did this for a variation on the Ellsberg paradox II. They considered problems in the form of Ellsberg's paradox II with the additional information that there are at least x yellow balls and at least y black balls. Hence the outer and inner measure of the event {a yellow ball drawn}, {a black ball drawn} were varied. Therefore if $\mu^*(D) > \mu_*(D)$ they suggested that

$$P(D) = \frac{\mu^*(D) - \mu_*(D)}{2}$$

and the "ambiguity factor" is a function of the magnitude of the difference $\mu^*(D) - \mu_*(D)$.

It is not obvious how this result can be generalized. For example, assume that there exists an urn with the following contents: 90 balls are a mixture of yellow, black, and orange balls, 50 balls are a mixture of orange and green balls. We are also told that there are exactly 40 orange and yellow balls. There is no obvious way to determine the "most likely" number of orange balls (event (O)) and neither is the estimate

$$P(O) = \frac{\mu^*(O) - \mu_*(O)}{2}$$

an obvious conclusion.

Fellner's (1961) approach is similar to the approach developed here in the sense that the P-measure may not necessarily be additive. Thus, the difficulty arises as to how P-measure can be used to define an expected value. Fellner did not specify any properties he would expect the P-measure to have except that it must be transformable into a probability measure.

For example, if P_1 and P_2 are the P-measures of events B and $\overline{B}$, we form the corrected probabilities

$$\hat{P}_1 = P_1/(P_1 + P_2) \text{ and } \hat{P}_2 = P_2/(P_1 + P_2).$$

Hence $\hat{P}_1$ and $\hat{P}_2$ are additive for the events B and $\overline{B}$. However let $C \varepsilon \Theta$ such that $\mu(C) = P_1(B)$, and consider the alternatives:

$$X(\omega,a) = \begin{cases} \$100 & \omega \varepsilon B \\ \$0 & \omega \varepsilon B \end{cases} \qquad X(\omega,b) = \begin{cases} \$100 & \omega \varepsilon C \\ \$0 & \omega \varepsilon C \end{cases}$$

If $U(\$0) = 0$ then

$$U(100)\mu(C) = U(100)P_1(B).$$

Now if we use the $\hat{P}_1$, we must modify the utility function so that

$$U(100)\mu(C) = U_2(100)\hat{P}_1$$

that is, $U_2(100) = U(100)P_1/\hat{P}_1$.

Fellner therefore argues that $U_2(\cdot)$ contains not only the amount of money we may receive but also our "dislike" or "like" for the "ambiguity" of the event B. In this approach therefore we must first derive the P-measures such that they can be transformed to probabilities and secondly we must derive a utility function for each action. Therefore rather than having an "ambiguity factor" depending on the action as in Ellsberg's case, Fellner suggested a different utility function for each action.

7.0 Derivation of P-measure

In section 6 we summarized our set of assumptions and axioms for a theory of decision making under partial risk. Although those conditions are sufficient for the "expected utility" (Theorem 6.0.1) to hold, some difficulties arise when we attempt to derive the P-measure for arbitrary sets in β, unless additional assumptions are made. This is in contrast to Savage's approach. For example, if we derive the subjective probabilities using Axiom IV (or Savage Axiom S4) for a class of pairwise disjoint sets $B_1,\ldots,B_n$ then we know the subjective probability of all sets which can be formed by taking unions, intersections or complements of the sets $B_1,\ldots,B_n$. However, since the P-measure is not necessarily additive, that is, $P(B_1 \cup B_2)$ may not equal $P(B_1) + P(B_2)$, these probabilities are generally not known. Hence to derive the P-measure for all possible events, we must consider every possible combination of union, intersection and complement. This would be very time consuming if not impossible when the number of events is very large. However, if there exist some events for which the P-measure would have the standard properties of a measure, i.e., additivity, the derivation would be simplified substantially.

In section 7.1 we shall suggest those sets for which it seems most likely that the P-measure is additive. In section 7.2 we shall show that there always exists a P-measure having the properties we assumed. Finally, in section 7.3 we shall give some possible P-measures when β contains only finitely many sets.

7.1 Additivity of the P-measure

Before stating the axiom, we shall consider some of the differences between decision making under partial risk and risk. Ellsberg (1961) reported a variety of choices among hypothetical lotteries, implying that "ambiguity" associated with the probabilities of some events influences the choices. By "ambiguity" he meant that the probabilities were not specified precisely, defining it as

> "...a quality depending on the amount, type, reliability and 'unanimity' of information and giving rise to one's degree of 'confidence' in an estimate of relative likelihood". (p.657)

Becker and Brownson (1964) modified this definition to state

> "...ambiguity is defined by any distribution of probabilities other than point estimates". (p.64)

Implicit in both definitions is the existence of a "second order" probability distribution. However, if it were reasonable to assume a given second order probability distribution, then Marschak's Axiom M2 (see Appendix I) ought to be used to calculate the expected utility. In our development the existence of a probability measure is not necessary. Neither of these definitions is, therefore, appropriate here. Note, however, that in both definitions we may speak of the "ambiguity of the event C" if the probability of C is not precisely specified or derivable in the sense that the outer measure of C is strictly greater than the inner measure, i.e.,

$$\mu^*(C) > \mu_*(C).$$

Also, it seems obvious that some events may have "more" ambiguity than others. For example, if we know that an urn contains 100 balls which are a mixture of red and black, then the event of drawing a red ball has "more" ambiguity in this case than if we knew that there were between 49 and 51 red balls in the urn. We can, therefore, also speak of the "degree of ambiguity". We shall formalize both of these ideas in the following definition.

Definition 7.1.1. The degree of ambiguity of an event C is defined to be the difference between its outer and inner measures, i.e.,

$$\mu^*(C) - \mu_*(C).$$

We speak, therefore, only of "the ambiguity of a set C" if its "degree of ambiguity" is strictly greater than zero.

This definition has several implications concerning the degree of ambiguity of the events. We shall summarize some of these here. We shall state these implications as lemmas although the proofs are trivial and follow directly from the properties of inner and outer measure.

The first property is that an event and its complement must have the same degree of ambiguity.

Lemma 7.1.1. For any C

$$\mu^*(C) - \mu_*(C) = \mu^*(\overline{C}) - \mu_*(\overline{C}).$$

Proof. From Appendix II we have

$$\mu^*(C) = 1-\mu_*(\overline{C}) \qquad \text{for any event } C.$$

Therefore

$$\mu^*(C) - \mu_*(C) = [1-\mu_*(\overline{C})] - [1-\mu^*(\overline{C})]$$

$$= \mu^*(\overline{C}) - \mu_*(\overline{C}).$$

The other properties we shall consider are concerned with "the degree of ambiguity" of union of events.

If D is any event for which the probability is known, and C is any event disjoint with D, then it seems reasonable that the event $D \cup C$ has the same degree of ambiguity as C. This property is the next lemma.

Lemma 7.1.2. If $D \in \Theta$, $C \in \beta$ and $D \cap C = \emptyset$ then

$$\mu^*(C) - \mu_*(C) = \mu^*(D \cup C) - \mu_*(D \cup C).$$

Proof. By the properties of inner and outer measure we have

$$\mu^*(C \cup D) = \mu^*(C) + \mu^*(D)$$

$$\mu_*(C \cup D) = \mu_*(C) + \mu_*(D).$$

Hence

$$\mu^*(D \cup C) - \mu_*(D \cup C) =$$

$$= \mu^*(D) + \mu^*(C) - \mu_*(D) - \mu_*(C)$$

$$= \mu^*(C) - \mu_*(C)$$

since $\mu^*(D) = \mu_*(D)$.

The second additivity property is concerned with the occasion when both sets have a degree of ambiguity. Consider, for example, the union of an event C and its complement. In this case the degree of ambiguity must cancel out since the event $C \cup \bar{C}$ must have the degree of ambiguity of zero. In general nothing can be said of the union of events where each has a positive degree of ambiguity.

Lemma 7.1.3. If $C, D \in \beta$ and there exist two sets $E, F \in \Theta$ such that $C \subset E$, $D \subset F$ and $E \cap F = \emptyset$ then

$$\mu^*(C \cup D) - \mu_*(C \cup D) = (\mu^*(C) - \mu_*(C)) + (\mu^*(D) - \mu_*(D)).$$

Proof. The two sets E and F give the condition for the degree of ambiguity of the two sets C and D to be separated and in this case the proof follows directly from the equations

$$\mu^*(C \cup D) = \mu^*(C) + \mu^*(D)$$

and $$\mu_*(C \cup D) = \mu_*(C) + \mu_*(D).$$

Here we shall assume that the P-measure is dependent on the degree of ambiguity. We shall illustrate what we mean by this in the following examples. First we consider some of the additivity properties using Ellsberg's paradox II, and a modification thereof. Consider the alternatives:

b) receiving \$100 is a yellow ball is drawn
   receiving \$0 otherwise

c) receiving \$100 is a red or a black ball is drawn
   receiving \$0 otherwise.

If we assume the decision maker has indicated his beliefs are such that $P(B) = P(Y)$ then it is implied that he believes the best estimator for the probability of drawing a black ball is equal to 1/3. Due to the degree of ambiguity, however, he "discounts" that probability to $P(Y)$ (where $P(Y)$ may actually be greater than 1/3).

The amount that he discounts the event is equal to $1/3 - P(B)$. Therefore, since the event $R \cup B$ of alternative c must have the same degree of ambiguity by lemma 7.1.2, he would "discount" the event by the same amount. That is,

$$\begin{aligned} P(R \cup B) &= 2/3 - (1/3 - P(Y)) \\ &= 1/3 + P(Y) \\ &= 1/3 + P(B). \end{aligned}$$

This implies therefore that the P-measure is additive for the event $R \cup B$. This would also imply that if $P(R) \geqslant P(Y)$, $P(Y) = P(B)$ then $P(R \cup B) \leqslant P(Y \cup B)$ which would explain Ellsberg's Paradox II. We shall also consider the case where neither the sets C or D belong to Θ; this decision problem is a variation of Ellsberg's paradox.

Modified Ellsberg's paradox II. Consider an urn which contains 100 balls, 50 of which are a mixture of red and yellow balls and 50 of which are a mixture of green and orange balls. Let us assume that the event $\{R \cup Y\}$ = {a red or yellow ball drawn} has a probability of 1/2 of occurring. Similarly the probability of the event $\{G \cup O\}$ = {a green or an orange ball drawn} is 1/2. The "degree of ambiguity" for the event $\{R\}$ = {a red ball is drawn} is, therefore, equal to 1/2 since μ^* {a red ball is drawn} = 1/2 and μ_* {a red ball is drawn} = 0. Similarly the degree of ambiguity for the event $\{G\}$ = {a green ball is drawn} is equal to 1/2. If we now consider the event $\{R \cup G\}$ = {a red or a green ball is drawn}, we note that the degree of ambiguity is equal to 1, i.e., the sum of the "degree of ambiguity" of the events. Therefore if $P(R) = P(Y)$ and $P(G) = P(O)$ that is, the event $\{R\}$ has been "discounted" by $1/4 - P(Y)$, and similarly $\{O\}$ has been "discounted" by $1/4 - P(G)$. Since the degree of ambiguity for the event $\{R \cup O\}$ is the sum of the degrees of ambiguity of both events, it seems reasonable to subtract the discount factors from both of these.

That is,

$$P(R \cup O) = 1/2 - (1/4 - P(R)) - (1/4 - P(O))$$

$$= P(R) + P(O).$$

Hence the probability for the event $\{R \cup O\}$ is additive for the event $\{R\}$ and $\{O\}$. We summarize these ideas now as Axiom VII.

Axiom VII. For each $C \varepsilon \beta$ and $D \varepsilon \Theta$ we have

$$P(C) = P(C \cap D) + P(C \cap \overline{D}).$$

If in Ellsberg's paradox II we let $C = (R \cup B)$ and $(D) = (R)$ then Axiom VII implies that

$$P(R \cup B) = P(R) + P(B) \text{ if } R \cap B = \emptyset$$

which we illustrated in the beginning of this section. Similarly, if in the modification we let $C = (R \cup O)$ and $D = (R \cup Y)$ then

$$P(R \cup O) = P(R) + P(O),$$

as illustrated in the last example. Axiom VII therefore specifies the sets for which the P-measure ought to be additive.

7.2 Existence of P-measure

In section 3.2 we noted the P-measure was "nearly" additive for certain sets; in Axiom VII we assumed additivity for exactly those sets. Hence at least we are consistent in the formulations of the axioms.

One of the difficulties which arises when specifying a set of axioms we believe the decision maker ought to have, is that there may not exist a P-measure with those properties. For example, we have shown that the Savage axioms contradict Axiom V. In section 8 we shall see that if we require the P-measure to be a probability, additional restrictions must be imposed on β. That is, β can not be an arbitrary set of events and hence A can not be an arbitrary set of actions. The question now arises as to whether the axioms we have assumed would restrict β in any way. If we translate the axioms into properties the P-measure must satisfy, then we can summarize the problem as follows:

Given an arbitrary probability space (Ω,Θ,μ) and an arbitrary σ-algebra β containing Θ, does there exist a set function P such that

1) $P(C)$ is defined for all $C \varepsilon \beta$
2) $P(C) = \mu(C)$ for all $C \varepsilon \Theta$
3) $P(C) \leqslant P(D)$ for all $C \subset D$ and $C,D \varepsilon \beta$
4) If $B_i \subset B_{i+1}$ and $B_i \varepsilon \beta$ for all i, and $\lim B_i = B$, then $\lim P(B_i) = P(B)$
5) $P(C) = P(C \cap D) + P(C \cap \overline{D})$ for all $C \varepsilon \beta$ and $D \varepsilon \Theta$.

Fortunately this problem is easily solved since μ^* has these properties. However 4) is not always true for an arbitrary outer measure, but only a regular outer measure (see Appendix II). This does not concern our development here, however, since the outer measure we are considering is always induced by the measure μ and hence therefore always regular.

7.3 Some possible P-measures

In this section we shall consider some plausible P-measures when β contains only a finite number of sets. Therefore we are only interested in P-measures which satisfy properties 1,2,3 and 5 in section 6.2.

For simplicity we shall assume that there exist finitely many disjoint subsets of Ω denoted by $D_1,\ldots,D_n$ which generate the σ-algebra Θ. Let C be any set such that $C \notin \Theta$, and let β be the σ-algebra generated by $C,D_1,\ldots,D_n$.

We also assume that a measure μ is defined on Θ. Our object is then to define a P on β.

We note that both $P(\cdot) = \mu^*(\cdot)$ or $P(\cdot) = \mu_*(\cdot)$ satisfy all properties 1 to 5, in this case. If we accept μ^* as our P-measure then, from a decision maker's point of view this would imply that in the Ellsberg paradox, $P(B) = 2/3$ and $P(Y) = 2/3$. If we accept μ_* as our P-measure, $P(B) = 0$ and $P(Y) = 0$. Both seem very unlikely to be accepted and empirical studies indicate that they are not accepted.

A more likely candidate would be a weighted average of μ^* and μ_*, i.e., $P_1(C) = \alpha\mu^*(C) + (1-\alpha)\mu_*(C)$ for all $C \in \beta$, where $\alpha \in (0,1)$. It is obvious that P_1 satisfies the first of the properties on p.181 and similarly the second follows since $\mu^*(C) = \mu_*(C) = \mu(C)$ for all $C \in \Theta$. The third property follows since μ^* and μ_* are monotone. The fourth follows obviously, and hence we need only show that the fifth property holds.

Let D be an arbitrary set belonging to Θ and $C \in \beta$. Then we have

$$\mu^*(C) = \mu^*(C \cap D) + \mu^*(C \cap \bar{D})$$

and $$\mu_*(C) = \mu_*(C \cap D) + \mu_*(C \cap \bar{D}) \quad \text{(see Appendix II)}$$

Therefore

$$\begin{aligned}
P_1(C) &= \alpha\mu^*(C) + (1-\alpha)\mu_*(C) \\
&= \alpha[\mu^*(C \cap D) + \mu^*(C \cap \bar{D})] + \\
&\quad + (1-\alpha)[\mu_*(C \cap D) + \mu_*(C \cap \bar{D})] \\
&= [\alpha\mu^*(C \cap D) + (1-\alpha)\mu_*(C \cap D)] + \\
&\quad + [\alpha\mu^*(C \cap \bar{D}) + (1-\alpha)\mu_*(C \cap \bar{D})] \\
&= P_1(C \cap D) + P_1(C \cap \bar{D}),
\end{aligned}$$

and hence the fifth property is satisfied by $P_1(\cdot)$.

In terms of the Ellsberg paradox II, this would imply that $P(B) = P(Y)$ and may be any number between 0 and 2/3 inclusive and $P(B)+P(Y) = P(B \cup Y)$ if and only if $P(B) = 1/3$. Therefore, using this measure, Ellsberg's paradox is easily explained. A preference of a bet on {R} over {B} indicates $P(R) = 1/3 > P(B)$ and since $P_1(B) = P_1(Y)$ we have

$$P_1(B \cup Y) = 2/3$$

$$P_1(R \cup Y) = P_1(R) + P_1(Y) \quad \text{by Axiom VII}$$

$$= 1/3 + P_1(Y)$$

$$< 2/3 \quad \text{since } P_1(Y) < 1/3.$$

Therefore a preference of a bet on $\{R\}$ over $\{B\}$ implies a preference of a bet on $\{B \cup Y\}$ over $\{R \cup Y\}$. Similarly a preference of $\{B\}$ over $\{R\}$ would imply a preference of $\{R \cup Y\}$ over $\{B \cup Y\}$, and indifference between $\{R\}$ and $\{B\}$ would indicate an indifference between $\{B \cup Y\}$ and $\{R \cup Y\}$.

This definition of P-measure can be generalized to the case where β contains a finite collection of non-measurable sets. That is we assume that there exists a collection of disjoint subsets of Ω, denoted by $D_1,\ldots,D_m$ and let Θ be the σ-algebra generated by the sets $D_i, i=1,\ldots,m$. Let $C_1,\ldots,C_n$ be subsets of Ω such that $C_i \not\in \Theta$ $i=1,\ldots,n$, and let β be the σ-algebra generated by $C_1,\ldots,C_n$, $D_1,\ldots,D_m$ and again we assume that a measure μ is defined on Θ.

As before we shall extend μ to β. One way of doing so would be to define $E_{ij} = D_i \cap C_j$, $i=1,\ldots,m$, $j=1,\ldots,n$, and let $E_{ij} \neq \emptyset$, $j \in T^i$, $E_{ij} = \emptyset$, otherwise. That is, T^i is an index set for which the set E_{ij} is not equal to $\emptyset$ for a fixed i.

Define $P_2(E_{ij}) = \alpha_i \mu^*(E_{ij}) + (1-\alpha_i)\mu_*(E_{ij})$ where α_i is a real-valued function on the cardinality of T^i to the interval [0,1]. First consider the union $\bigcup_{j\in S} E_{ij}$, where $S \subset T^i$

Define

$$P_2(\bigcup_j E_{ij}) = \begin{cases} P_2(D_i) & T^i \subset S \\ \sum_{j\varepsilon s} P_2(E_{ij}) & \text{otherwise.} \end{cases}$$

Finally, if F is an arbitrary set in β let G be the largest subset of F, such that $G \varepsilon \Theta$, we define

$$P_2(F) = P_2(G) + \sum_{i=1}^{m} P_2(D_i \cap (F-G)).$$

Clearly, the first four properties hold for the same reason as for P_1. To show that the fifth property holds, let F and B be arbitrary sets belonging to β and Θ respectively. Then we are required to show that

$$P_2(F) = P_2(F \cap B) + P_2(F \cap \overline{B}).$$

Let G be the largest subset of F such that $G \varepsilon \Theta$, and then, clearly, $B \cap G$ must be the largest subset of $F \cap B$ such that $B \cap G \varepsilon \Theta$. Similarly, $\overline{B} \cap G$ must be the largest subset of $F \cap \overline{B}$.

$$P_2(F) = P_2(G) + \sum_{i=1}^{m} P_2(D_i \cap (F-G))$$

$$= P_2(G \cap B) + P_2(G \cap \overline{B}) + \sum_{i=1}^{m} P_2(D_i \cap (F-G)).$$

Since D_i either is contained in B or $\overline{B}$,

$$P_2(D_i \cap (F-G)) = P_2(D_i \cap ((F \cap B) - (G \cap B)) +$$

$$P_2(D_i \cap ((F \cap \overline{B}) - (G \cap \overline{B})).$$

Hence by recombining terms

$$P_2(F) = P_2(F \cap B) + P_2(F \cap \overline{B}).$$

Applying this P-measure to the Ellsberg paradox II, we would obtain exactly the same measure as discussed before. However, for the modified paradox we would have $P_2(G \cup W) = P_2(G \cup O) = P_2(W \cup O) = P_2(G) + P_2(W) = P_2(G) + P_2(O) = P_2(W) + P_2(O)$ and no relation would necessarily hold between $P_2(B)$ to $P_2(G)$. The preferences between {B} and {G} may be either way without contradicting the theory.

Note that the way we constructed the last P-measure works for the case where β is finite. That is, first define P for the "smallest" unit E_{ij}, next define the P-measure for sets $\bigcup E_{ij} \subset D_i$, and finally impose the additivity condition,

$$P(F) = P(G) + \sum P(D_i \cap (F-G)).$$

In doing so, we would always satisfy the five properties on p. 186.

The P-measure for an arbitrary set E_{ij} can probably be approximated in general as

$$P(E_{ij})=(\alpha_i\mu^*(E_{ij}) + (1-\alpha)\mu_*(E_{ij}))e^{k_i(\mu^*(E_{ij}) - \mu_*(E_{ij}))}$$

where α_i = 1/cardinality of T^i and k_i any real number.

To explain this formula we note that for the Ellsberg paradox II $\alpha_i\mu^*(E_{ij}) + (1-\alpha_i)\mu_*(E_{ij})$ for the event {B} and {Y} would be 1/2(2/3) + 1/2(0) = 1/3.

For the modified paradox and for the event {B} and {Y}, we would have 1/2(1/3) + 1/2(0) = 1/6 and for the event {G}, {W}, and {0} we would have 1/3(1/2) + 1/3(0) = 1/6.

Hence if $k_i = 0$, we would have the P-measure as a measure (i.e., additive). If $k_i < 0$, this implies we would discount this measure based on the difference between the outer and inner measure. If $k_i > 0$, this would indicate a preference of uncertainty which increases as the difference between outer and inner measures increases.

None of the P-measures mentioned here are, of course, necessary for the theory to hold. We have suggested some of these as they seem to have empirical support, and would explain the difficulties in the paradoxes concerning uncertainty vs. risk. The advantage of approximating the P-measure is, of course, that we only need to derive one constant k_i and from this we can derive the P-measure for all sets.

8.0 The P-measure as a probability measure

In decision theory one frequently comes in contact with other names used for a probability measure. The most common ones, in alphabetic order, are:

Degree of confirmation	Mathematical probability
Degree of conviction	Objective probability
Degree of rational belief	Personal probability
Empirical probability	Physical probability
Geometric probability	Psychological probability
Impersonal probability	Random chance
Inductive probability	Relative frequency
Intuitive probability	Statistical probability
Judgment probability	Subjective probability
Logical probability	

(Fishburn, 1964, pp.132)

In Bayesian statistics we also have the additional terms of prior probability and posterior probability. Most of these satisfy Kolmogorov's axiom (see Appendix II); others differ only by assuming finite additivity rather than σ-additivity. For our purpose we shall divide them into two categories. The first category assumes σ-additivity, and will be called mathematical probabilities; the second category assumes only finite additivity and will be called Savage probabilities.

We are concerned only with the properties of probabilities. Specifically, our interest is in the implications of assuming the P-measure to be finitely additive or equivalently σ-additive. In section 2.0 we assumed that there exists a probability space (Ω,Θ,μ). We also assumed that the reward functions $X(\cdot,a)$, $a \in A$ are not all measurable, that is Θ is too small. We had therefore to extend the measure μ to the smallest σ-algebra for which all functions $X(\cdot,a)$ are measurable. It would, therefore, be of interest to determine the largest σ-algebra for which μ can be extended as a probability measure. In section 8.1, we shall consider these arguments for Savage's probabilities. In section 8.2 we shall consider the P-measure as a σ-additive measure and discuss the possible extension of μ. In section 8.3 we shall consider an alternative to decision making under partial risk. The reason for doing so here is that this alternative also induces a probability measure.

8.1 Savage extensions

It is well known, and proven in most basic textbooks on measure theory, that if (Ω,Θ,μ) is defined such that Ω is the real line, Θ all Borel sets, and μ the Lesbeque measure, it is impossible to extend μ to all subsets of the real-line if it is also required that μ be σ-additive. It is also well known that if we only require a finitely additive measure the extension exists (see Roydėn, 1968, p.53). Most mathematicians today assume σ-additivity of the measure although some research is still taking place concerning finitely additive measures. From a decision maker's viewpoint the implication of the measure not being defined on all subsets can be illustrated by the following simple situation. There exists a set $D \subset \Omega$ such that if we are offered a lottery ticket defined by

$$X(\omega,a) = \begin{cases} \$100 & \omega \in D \\ \$0 & \omega \in \overline{D} \end{cases}$$

it would be impossible to determine its equivalent value, i.e., how much it is worth.

Savage, for one, disliked this implication, and was therefore forced to assume only additive measures. In the Savage approach a preference ordering was determined on the subsets of Ω. This can be done by a variation of Axiom III. One method determines a preference ordering on 2^{Ω} by letting $B \subset \Omega$, $D \subset \Omega$, $U(r) > U(s)$ and

$$X(\omega,a) = \begin{cases} r & \omega \in B \\ s & \omega \in \overline{B} \end{cases} \qquad X(\omega,b) = \begin{cases} r & \omega \in D \\ s & \omega \in \overline{D}. \end{cases}$$

Then $a \overset{A}{\geqslant} b$ implies $B \overset{2^{\Omega}}{\geqslant} D$.

Several authors (De Finetti, 1937; Savage, 1954) have studied the problem of determining conditions on the ordering on 2^{Ω} under which there exists a real-valued order preserving function P on 2^{Ω} that can be interpreted as an additive probability measure. The extension to σ-additivity was made by Kraft, Pratt and Seidenberg (1959). In section 6.2 of Part I we listed the conditions on this preference ordering for the existence of a probability measure such that if $A \overset{2^{\Omega}}{\leqslant} B$ then $P(A) \leqslant P(B)$. Although a P-measure is not necessarily a probability measure, it would be of interest to determine what conditions on the preference ordering we satisfied using our P-measure.

We shall repeat for easy reference, the axiom stated in section 6.2 of Part I.

<u>Axiom of ordering</u>

1. If $C \in \Theta$, $B \in \Theta$, then either $C \overset{2^{\Omega}}{\leqslant} B$ or $B \overset{2^{\Omega}}{\leqslant} C$.
2. For any set $C \in \Theta$, $C \overset{2^{\Omega}}{\leqslant} C$.
3. If $C \overset{2^{\Omega}}{\leqslant} B$ and $B \overset{2^{\Omega}}{\leqslant} D$ then $C \overset{2^{\Omega}}{\leqslant} D$.
4. $\emptyset \overset{2^{\Omega}}{<} \Omega$ and for any event $C, \emptyset \overset{2^{\Omega}}{\leqslant} C \overset{2^{\Omega}}{\leqslant} \Omega$.

All these axioms can easily be derived from Axioms I and II.

Axiom of monotonicity

1. If $B_1 \cap B_2 = \emptyset$, $C_1 \overset{2^\Omega}{\leqslant} B_1$ and $C_2 \overset{2^\Omega}{\leqslant} B_2$,

   then $C_1 \cup C_2 \overset{2^\Omega}{\leqslant} B_1 \cup B_2$.

2. If $B_1 \cap B_2 = \emptyset$, $C_1 \overset{2^\Omega}{\leqslant} B_1$ and $C_2 \overset{2^\Omega}{<} B_2$,

   then $C_1 \cup C_2 \overset{2^\Omega}{<} B_1 \cup B_2$.

The axioms of monotonicity may not be satisfied by the P-measure. Consider, for example, Ellsberg's paradox II.

C_1 = {a yellow ball drawn}

B_1 = {a red ball drawn}

C_2 = {a black ball drawn}

B_2 = {a yellow ball drawn}

Then we might have the preference $B_1 \cap B_2 = \emptyset$, $C_1 \overset{2^\Omega}{\leqslant} B_1$, $C_2 \overset{2^\Omega}{\leqslant} B_2$ but

$$C_1 \cup C_2 \overset{2^\Omega}{\geqslant} B_1 \cup B_2.$$

Axiom of monotone sequence

For every monotone sequence increasing events such that $C_n \nearrow C$ and an event B such that

$$C_n \overset{2^\Omega}{\leqslant} B, \text{ for all } n, \text{ then } C \leqslant B.$$

This axiom follows from Axiom V and Axiom III.

To see this, define a sequence of reward functions as follows:

$$X(\omega, a_n) = \begin{cases} r & \omega \in C_n \\ s & \omega \in \overline{C}_n \end{cases}$$

where $U(r) \geq U(s)$ and $C_n \in \beta$. By Assumption 3, $a_n \in A$ for all n. Since C_n is a sequence of increasing sets converging to C, a_n converges to $a \in A$ where a is defined by the reward function

$$X(\omega, a) = \begin{cases} r & \omega \in C \\ s & \omega \in \overline{C} \end{cases} .$$

If $b \in A$, and b is defined by the reward function

$$X(\omega, b) = \begin{cases} r & \omega \in B \\ s & \omega \in \overline{B} \end{cases}$$

where $B \in \beta$. By lemma 3.2.1

$$P(C_1) \leqslant P(C_2) \leqslant P(C_3) \leqslant \ \ldots$$

and $\lim P(C_n) = P(C)$. Therefore if $P(C_n) \leqslant P(B)$ for all n, then $P(C) \leqslant P(B)$ also.

Axiom of partition of event

Every event can be partitioned into two equally probable events. This assumption is needed to specify the probability measure. Villegas (1964) showed that this axiom, under certain assumptions, is equivalent to the existence of a random variable with uniform distribution (Assumption 3). Therefore, the only axiom which may not hold in the P-measure approach is the axiom of monotonicity.

8.2 Mathematical probabilities

In this section we shall also assume there exists a probability space (Ω,Θ,μ). Our aim is to study the possible extension of Θ to a σ-algebra containing Θ. The Savage probability satisfies:

1) $\bar{\mu}(C)$ exists for all $C \subset \Omega$.

2) $\bar{\mu}(C) = \mu(E)$ $C \in \Theta$.

3) If $C_1,\ldots,C_n$ are disjoint sets then

$$\bar{\mu}\left(\bigcup_{i=1}^{n} C_i\right) = \sum_{i=1}^{n} \bar{\mu}(C_i).$$

In this section we shall replace 1) by:

1) $\bar{\mu}(C)$ exists for all $C \in \beta$ where β is a very special σ-algebra,

and replace 3) by:

3) If C_i $i=1,\ldots$ is a disjoint sequence $C_i \in \beta$, then

$$\bar{\mu}\left(\bigcup_{i=1}^{\infty} C_i\right) = \sum_{i=1}^{\infty} \bar{\mu}(C_i).$$

The first extension result we shall state is the standard textbook case which can be justified in the following way. Assume that for a given set B we know the probability of occurrence is equal to one-half. Let D be a subset of B which does not belong to Θ. If we wish to include D in the σ-algebra the difficulty is to determine which measure to assign D such that we do not contradict any of the axioms of the measure.

There is one case, however, when this problem does not arise; that is, if $\mu(B) = 0$. Then, for any subset the only possible measure we could assign would be zero. When this is done for all subsets of the measure zero, it is called the completion of the probability space.

Definition. A measure μ on a σ-algebra β is said to be complete if and only if whenever $D \subset B$ and $\mu(B) = 0$, $B \varepsilon \Theta$ then $D \varepsilon \beta$.

The completion of a probability space (Ω,Θ,μ) is defined as follows. Let β be the σ-algebra generated by the set $\{G \cup D\}$ where $G \varepsilon \Theta$ and $D \subset B$ where $\mu(B) = 0$ for some $B \varepsilon \Theta$. We extend μ to $\bar{\mu}$ by $\bar{\mu}(G \cup D) = \mu(G)$. The implication is that if a decision maker refuses to pay anything for a lottery ticket which wins any amount if B occurs, he would also refuse to pay anything for a lottery ticket which wins the same amount if any subset of B occurs. This is an obvious result from a decision maker's point of view and hence we can always assume that the probability space we are working with is complete.

There also exists a second case when this problem is easily solved. If we again assume $D \subset B$, then by monotonicity of the measure we know that $\mu(D)$ must be less than $\mu(B)$. If we also know that there exists a set N contained in $D(N \subset D)$ such that $N \varepsilon \Theta$ and $\mu(N) = \mu(B)$, then clearly $\mu(D) = \mu(B)$. This method was used by Lebesque (1901) in defining a measurable set for which the inner and outer measures are equal.

It is also known that the σ-algebra generated by all sets such that inner and outer measures are equal is the largest

σ-algebra for which μ can be extended uniquely, we shall denote this σ-algebra by $\bar{S}$. The outer measure will be denoted as before by μ^* and the inner measure as before by μ_*.

If we now remove the condition that μ be uniquely extended, can we still extend $\bar{S}$? Let Z be any set such that $Z \not\in \bar{S}$, then we denote the smallest σ-algebra generated by the collection of sets Z and all sets in $\bar{S}$ by $(\bar{S},Z)$. We shall next state how μ can be extended to this σ-algebra.

Theorem 8.2.1. Los and Marczewski (1949)

Let μ be a probability on a σ-algebra $\bar{S}$ on Ω and $Z \subset \Omega$. Define $\bar{\mu}$ and $\underline{\mu}$ for each set $E \in (\bar{S},Z)$ by

$$\underline{\mu}(E) = \mu_*(E \cap Z) + \mu^*(E \cap \bar{Z})$$

$$\bar{\mu}(E) = \mu^*(E \cap Z) + \mu_*(E \cap \bar{Z})$$

Then $\underline{\mu}$ and $\bar{\mu}$ are extensions of μ to $(\bar{S},Z)$ and $\underline{\mu}(Z) = \mu_*(Z)$ and $\bar{\mu}(Z) = \mu^*(Z)$.

Corollary 8.2.2. The function

$$m(E) = (1-\alpha)\underline{\mu}(E) + \alpha\bar{\mu}(E) \qquad \alpha \in (0,1)$$

is also an extension of μ to $(\bar{S},Z)$.

We note that if $\mu^*(Z) = \mu_*(Z)$ then $\bar{\mu}(E) = \underline{\mu}(E)$ for all $E \in (\bar{S},Z)$ so the only interesting case is if $\mu^*(Z) > \mu_*(Z)$.

We also note that for any set $E \in (\bar{S},Z)$ we have

$$\mu_*(E) \leqslant \underline{\mu}(E) \leqslant \mu^*(E)$$

$$\mu_*(E) \leqslant \overline{\mu}(E) \leqslant \mu^*(E)$$

$$\mu_*(E) \leqslant m(E) \leqslant \mu^*(E).$$

We can obviously repeat this process by considering the inner and outer measure generated by $\underline{\mu}$, $\overline{\mu}$ or m, and hence we can extend to the larger σ-algebra generated by adding a finite number of sets. Los and Marczewski (1949) showed, however, that we may not be able to add a countable number of sets. That is, there may still exist sets which are still not measurable, and therefore, we still have the difficulty of explaining the existence of a simple decision problem in section 8.2.

8.3 Probability measures on Fuzzy sets

The theory of decision making in a fuzzy environment (see Zadeh (1965), (1968), (1969)) is related to the decision making under partial risk. The difference however is that events in the fuzzy environment are not clearly defined. Examples of fuzzy sets are: "X is approximately equal to 5"; or "In twenty tosses of a coin there are several more heads than tails". Because of the vagueness of the description of events difficulties in specifying the probabilities for these events can be anticipated, as with decision making under partial risk. In this section we shall summarize how this problem is handled for fuzzy sets, and relate the idea to decision making under partial risk. To do so we must define fuzzy sets in more mathematical terms.

Definition. (Zadeh, 1965). Let $\Omega = \{w\}$ be a collection of objects (states). A fuzzy set B in Ω is a set of order pairs

$$B = (w, J_B(w)) \qquad w \in \Omega$$

where $J_B(w)$ is a real valued function from Ω to $M = [0,1]$.

If $M = \{0,1\}$ then this is equivalent to the normal definition of a subset of Ω.

The function $J_B(w)$ can be thought of as "the degree of confidence" we have of w belonging to B. The function is equal to 1 if we are certain of w belonging to B and 0 if we are certain that this is not the case.

To illustrate the function $J_B(w)$, let Ω be a set of 60 balls, numbered from 1 to 60. All the balls are either yellow or black and we know for certain that the first 10 balls are yellow and the last 10 are black. Somewhere between 11 and 51 there exists a number γ such that all numbers below including γ are yellow, and those above are black. The function $J_B(w)$ where B is the number of black balls can be illustrated graphically as in Fig. 7.4.

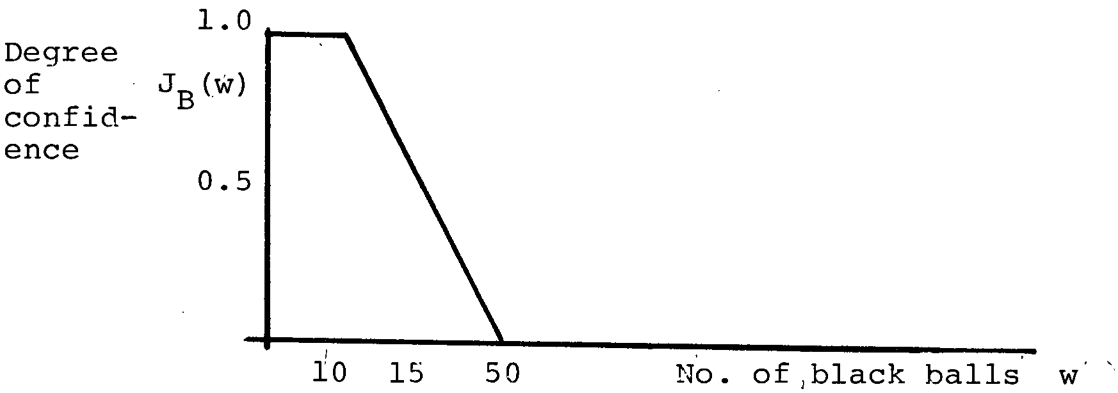


Fig. 7.4. Illustrating the degree of confidence in relation to the possible number of black balls in an urn with partial knowledge.

For example the graph illustrates that we are 50% confident that there are at least 15 black balls.

All standard set operations can be defined for fuzzy sets as follows:

The intersection of two sets C and B is denoted by $C \cap B$ and defined by the membership function

$$J_{C \cap B}(\cdot) = \min(J_C(\cdot), J_B(\cdot)).$$

The complement $\bar{C}$ of the set C is defined by the membership function

$$J_{\bar{C}}(\cdot) = 1-J_C(\cdot).$$

The union $C \cup B$, of the set C and B is defined as

$$J_{C \cup B}(\cdot) = \max(J_C, J_B).$$

For example the complement states that if we are 50% confident that there are at least 15 black balls we are also 50% confident there are at most 45 yellow balls.

In defining probabilities on fuzzy sets we assume first there exists a probability space (Ω, Θ, μ) where Θ does not contain fuzzy sets.

Hence

$$\mu(B) = \int J_B(w)\,d\mu$$

where $J_B(w)$ is the indicator function for the set B. Let β denote the σ-algebra of fuzzy sets then if $C \in \beta$ the probability of C is defined by

$$\bar{\mu}(C) = \int J_C(w)\,d\mu.$$

It can be shown that $\bar{\mu}$ satisfies the standard properties of a probability measure. Hence this approach is not different from any other approach which develops probabilities on the set β.

It is very likely this approach may be used to develop similar results as P-measures. However, in this case the standard definition of intersection, union and complement would have to be re-defined.

9.0 Empirical studies related to P-measure

The reason for considering a theory which differentiates between risk and uncertainty is that this seems to be a commonly made distinction. In this section we shall substantiate this claim by summarizing some empirical studies made concerning this point. We shall also show that these studies indicate a correspondence between the degree of ambiguity and P-measure.

Our first objective will be to show that if Axiom IVb is accepted, the P-measure defined thereby is not necessarily a probability measure. This can be done by using Ellsberg's paradox I. Recall that in Ellsberg's paradox I we considered two urns, one with 50 red and 50 black balls, and the second where the composition of red and black balls was unknown. Alternatives a and c referred to actions defined on balls drawn from the urn with the known composition and alternatives b and d on the other. If the P-measure P is a probability then a preference of a over b implies d must be preferred to c. Similarly if b is preferred to a, then c must be preferred to d. For a detailed study of the Ellsberg paradox I, see section 6.4, Part I, and references to other related studies.

Let us now summarize some results concerning this paradox from the study by MacCrimmon and Larsson (1975). Fifteen out of 19 subjects chose a P-measure which is not a probability measure. Another subject ranked three of the bets equally with alternative b lower. Two subjects ranked all four choices equally and the remaining subject ranked the choices in accordance with a probability measure.

Hence 15 out of 19 did not choose alternatives as though the P-measure were a probability measure and only three did choose as though it were. Although other studies have not shown such a large rejection of P-measure as a probability, the percentage of rejection has been substantial. This indicates, therefore, that it is reasonable to assume that there are differences between risk and uncertainty.

A study made by Becker and Brownson (1964) shows support for Axiom VII and shows that ambiguity can be measured as the difference between the outer and inner measure. Their study also indicates that the P-measure is not always a probability measure. Their study was based on a variation of Ellsberg paradox I. They considered five urns with the maximum and minimum number of red and black balls as shown below:

	Red Balls		Black Balls	
	Minimum Number	Maximum Number	Minimum Number	Maximum Number
Urn I	0	100	0	100
Urn II	50	50	50	50
Urn III	15	85	15	85
Urn IV	25	75	25	75
Urn V	40	60	40	60

The object of their study was to find support for the following hypotheses:

Hypothesis I.- Individuals are willing to pay money to avoid actions involving ambiguity.

The experiment sought to verify Ellsberg's findings under actual payoff situations. (Evidence supporting this hypothesis does no more than confirm that a third variable appears to be affecting decision behavior of some people; it does not lend differential support to any particular definition of the third variable.)

Hypothesis II.- Some people behave as if they associate ambiguity with the distribution on each probability.

(Becker and Brownson, p.65)

The result of their study was specified in two sets. First, it was determined whether the subjects had a preference as Ellsberg predicted by Ellsberg paradox I. The result was as follows:

Responses to Ellsberg questions	Number of Subjects
50 - 50 preferred	16
0 - 100 preferred	1
Indifferent	5
Color preference	12
Total	34

Hence they found a large percentage having a colour preference as well as a high percentage rejecting the P-measure as a probability measure. Fifteen out of the sixteen (one was omitted due to oversight on the part of the experimenter) were selected for further study.

Note that the inner measure of drawing a red ball is equal to the minimum number of red balls and the outer measure equal to the maximum number of red balls, and similarly for black balls. Therefore, for example, the degree of ambiguity is greater for the event of drawing a red ball from Urn I than for the event of drawing a red ball from Urn III.

In the second study the subjects were asked the amount they were willing to pay to have the ball drawn from their choice of urn for any two combinations if their winnings were \$1. Actual money was used for this experiment. For the exact findings of their results, we refer to Table 3 and Table 5 in their paper. However, their study clearly indicates that their subjects prefer the alternative with a smaller degree of ambiguity and they conclude that:

> "Evidence presented in Tables 3 and 5 confirms the first hypothesis that some people will pay to avoid an <u>ambiguous</u> course of action when that action has an expected value equal to an alternative <u>unambiguous</u> course of action. Fourteen of the fifteen subjects, all of whom indicated under non-payoff conditions that they had an aversion to ambiguity, were willing to pay a sum of money to have the opportunity to select their preferred course of action. When choosing between the 50-50 urn and the 0-100 urn, the subjects offered to pay an average of \$0.36 (the average of the amounts in the first column of Tables 3 and 5) in order to avoid an <u>ambiguous</u> course of action whose expected value was \$0.50. (One must wonder whether these subjects, in retrospect, would consider the discomfort avoided worth the price they paid.)
>
> The second hypothesis, that <u>ambiguity</u> is associated with the distribution on each probability, is also supported by the data. In all cases the preferred urn, for which the subject would pay a premium, was that urn which had the smaller range around $E(\overline{P}_r)$."
>
> (Becker and Brownson, p.70)

Their study was of course only designed for Ellsberg paradox I and does not in general indicate that most people prefer actions with no ambiguity, only in this very special example. Also note that in their statement

> "...an ambiguous course of action when that action has an <u>expected value</u> equal to an alternative unambiguous course of action...". (Becker and Brownson, p.70)

They imply there exists a second order distribution but this may not be the case.

The studies so far indicate that there are those which do differentiate between risk and uncertainty and that the degree of ambiguity may be one factor which determines the P-measure. However this does not imply that they accept the expected utility criterion being accepted in the first place; that is, the rejection may be by those who reject the expected utility criterion.

A study made by Yates and Zukowski (1975) indicates that many subjects who are consistent with the expected utility criterion for decision making under risk also differentiate between risk and uncertainty. They considered the following alternatives:

Alternative a.

Step 1: Designation of "valuable" chip. A color of poker chips, red or blue, is designated by the player as "valuable".

Step 2: Bookbag composition. Five valuable and five non-valuable chips are placed in a bookbag by the player.

Step 3: Drawing. The player draws one chip at random from the bookbag.

Step 4: Payoff. The player becomes entitled to receive $1. if a valuable chip is drawn in Step 3 or nothing if a non-valuable chip is drawn.

Alternative b.

Step 1: Designation of "valuable" chip. A color of poker chips, red or blue, is designated by the player as "valuable".

Step 2: First bookbag composition. Eleven white chips, marked 0 through 10, are placed by the player in the first of two bookbags.

Step 3: First drawing. The player draws one chip at random from the first bookbag.

Step 4: <u>Second bookbag composition</u>. The number of valuable chips corresponding to the number drawn in Step 3 are placed by the player in the second bookbag. Non-valuable chips are added to make a total of 10 valuable and non-valuable chips in the second bookbag. The chips are then mixed well.

Step 5: <u>Second drawing</u>. The player draws one chip at random from the second bookbag.

Step 6: <u>Payoff</u>. The player becomes entitled to receive $1. if a valuable chip is drawn in Step 5 or nothing if a non-valuable chip is drawn.

<u>Alternative c</u>.

Step 1: <u>Bookbag composition</u>. The player is informed that there are ten poker chips in a bookbag. Each poker chip is either red or blue. The number of chips of either color can be any number from zero through 10, with the total number of chips being 10. (The player is not informed of the number of chips of each color.)

Step 2: <u>Designation of "valuable" chips</u>. A color of poker chips, red or blue, is designated by the player as "valuable".

Step 3: <u>Drawing</u>. The player draws one chip at random from the bookbag.

Step 4: <u>Payoff</u>. The player becomes entitled to receive $1. if a valuable chip is drawn in Step 3 or nothing if a non-valuable chip is drawn.

We note that alternatives a and b are decision making under risk, and alternative c decision making under partial risk.

In Ellsberg's paradox I, it has sometimes been suggested that if the subject is asked from which urn he wants the red ball drawn, he may believe that the uncertainty alternative has been "rigged" against him, and therefore he may choose the one with given probabilities. Yates and Zukowski (1975) have eliminated this factor in their experiment by letting the player choose the colour. If the subjects do not have a colour preference then alternatives a and b ought to have the same preference if the expected utility criterion is used.

In their study they used 108 students and each student was classified in one of the following classes: a-b (20); b-a (14); a-c (19); c-a (12); b-c (16); c-b (27); i.e., 20 students had the choice between alternative a or alternative b (indifference not permitted) where alternative a was listed first; 14 students also had the choice between alternative a or alternative b, but in this case alternative b was listed first.

Actual payoff was used. Each student was required to choose which alternative he preferred in the class to which he belonged. Indifference was not permitted. Their result was as follows:

Alternative	Alternative a	b	c
a	-	16	7
b	18	-	14
c	24	29	-

where the column alternative was chosen over the row alternative.

Their study therefore shows that little difference exists between alternative a or alternative b, but alternative c is definitely not preferred to either a or b.

For the expected utility criterion to hold the expected frequency would be:

Alternative	Alternative a	b	c
a	-	17	21 1/2
b	17	-	15 1/2
c	15 1/2	21 1/2	-

Their result then indicates a substantial difference between alternatives c and a and also, to a smaller extent, between c and b. This indicates a distinction between risk and uncertainty.

These results can, therefore, be summarized as follows:

1) The P-measure derived based on Axiom IV may not be a probability. For Ellsberg's paradoxes there is a substantial percentage of rejection of the P-measure as a probability.

2) Ambiguity can be measured in terms of the difference between outer and inner measure.

3) Even if the expected utility criterion is accepted for decision making under risk, there is a large percentage of people who differentiate between risk and uncertainty.

10.0 Summary of Part II

In the second part of the thesis, we have suggested a criterion for decision making under partial risk. The definition of partial risk includes both decision making under risk and decision making under uncertainty. Therefore any criterion suggested for partial risk problems must also be accepted for risk and uncertainty problems. Here we assumed that the expected utility criterion is accepted for risk problems. Technically, therefore we have extended the expected utility criterion to decision making under partial risk. The idea of this extension is not new; both Savage (1954) and Fellner (1961) did similar extensions. We differ from Savage since his approach does not differentiate between risk and uncertainty. We also differ from Fellner's since his approach assumes a different utility function for each action.

What we have in common with each of these approaches is that for any extension of the expected utility criterion a probability measure must be derived on the set of possible rewards. It is sufficient that a monotonic continuous non-negative set function (called P-measure) is specified on the states of nature. Since every probability measure is a monotonic, continuous and non-negative set function, additional assumptions on the P-measure must be made for the P-measure to be a probability measure. Therefore, decision theorists who believe that the expected utility criterion ought to be used for decision making under uncertainty would also accept the approach developed here. They would perhaps insist that additional assumptions be made on the P-measure, forcing it to be additive.

We do not expect all decision theorists to support the assumptions developed here. We do, however, believe that if Axiom II is accepted, i.e., the expected utility criterion for decision making under risk, then the additional assumptions made here follow in the spirit of Axiom II and would be generally accepted.

In considering the axioms in more detail we expect to find that the most controversial axiom in Part II is Axiom V, which is where the sequence of reward functions is discussed. Here we have chosen to describe the axiom in terms of increasing reward functions. An equivalent result would be to describe the axiom in terms of decreasing reward functions. However if we state the axiom in terms of an arbitrary convergent sequence of reward functions, the existence of P-measure is not obvious (and has not yet been proven as far as we know). In terms of measure theory, Part II can be described as a mixture of standard measure theory and Savage's approach. One of the objectives of measure theory can be summarized by the following quote:

> "The length $l(I)$ of an interval I is defined, as usual, to be the difference of the endpoints of the interval ....In the case of length the domain is the collection of all intervals. We should like to extend the notion of length to more complicated sets than intervals.... we would like to construct a set function m which assigns to each set E in some collection Φ of sets of real numbers a nonnegative extended real number mE called the measure of E. Ideally, we should like m to have the following properties:
>
> i. mE is defined for each set E of real numbers; that is, $\Phi = T(R)$;
> ii. for an interval I, $mI = l(I)$;
> iii. if E_n is a sequence of disjoint sets (for which m is defined), $m(\bigcup_n E_n) = \sum_n m(E_n)$;

iv. m is translation invariant; that is, if E is a set for which m is defined and if $E + y$ is the set $\{x + y : x \in E\}$, obtained by replacing each point x in E by the point $x + y$, then

$$m(E + y) = mE.$$

Unfortunately, it is impossible to construct a set function having all four of these properties, and it is not known whether there is a set function satisfying the first three properties. Consequently, one of these properties must be weakened..."

(Royden, pp.52-53)

It is not obvious which one should be weakened. Mathematicians decided on the first of these, and restricted the total number of sets. The implication for decision theory is that there exists a subset D of Ω such that no price can be set on the lottery ticket which gives

$$X(\omega,a) = \begin{cases} \$100 & \omega \in D \\ \$0 & \omega \in \overline{D} \end{cases}.$$

I.e., there exist simple decision problems for which the expected utility criterion can not be used. This is an undesirable conclusion.

Savage (1954) rejected the third alternative, but only for a countable number of sets and as a consequence, he rejected Axiom V of Part II. We believe this to be an equally undesirable conclusion, and we also reject the third alternative, although we were more selective in determining on which sets the measure might be additive. This would, of course, give the desirable flexibility of distinguishing between risk and uncertainty.

Summary

In Part I of the thesis we discussed a set of axioms necessary for the expected utility criterion to hold. Empirical studies showed that by far the most controversial axiom was the additivity axiom (Axiom II, Part I) which specified the existence of a real-valued function h, such that if B_i $i=1,\ldots$ is a disjoint sequence of subsets of Ω, and each one belongs to β then $f(a) = \sum_i h(B_i, a)$.

This assumption obviously must be specified since the integral is σ-additive over disjoint sets. It also partly implies the existence of a measure W on β which reflects the decision maker's belief of the likelihood of a given event occurring. That is, if it is believed that B is more likely to occur than C, then $W(B) \geqslant W(C)$.

A second axiom which gives rise to some concern is Axiom IV. In this case a σ-algebra β containing Θ was assumed to exist. A probability measure μ was defined on Θ, and Axiom IV specified the relation between μ and W.

Several of the paradoxes discussed were shown to contradict either (or both) of these. That is a measure W on β could not be defined such that Axiom II would also be satisfied.

In Allais' paradox it was shown by MacCrimmon and Larsson (1975) that this contradiction to the axiom only occurred when the amount of reward money was substantially above what most people are used to handling. Hence, Allais' paradox does not contradict Axiom II when it pertains to decisions concerning amounts of money which most people are accustomed to handling. The same argument is not true, however, for Ellsberg's paradox, since in that situation contradictions occurred for even small amounts of money. Although both Ellsberg's paradox and Allais' paradox contradict Axiom II, it seems reasonable, however, to assume that Ellsberg's paradox differentiates between known versus unknown probabilities, rather than violating the additivity assumptions. These assumptions are, of course, not mutually exclusive since a rejection of the existence of probabilities on unknown events would also lead to the rejection of the additivity assumption.

In Part II of the thesis, we attempted to differentiate between unknown and known probabilities in an alternative manner. This was done by defining a less restricted set function on all events such that the set function is equal to the probability of known events but does not necessarily satisfy all the laws of probability for unknown events. That is, the set function is less structured in the latter case.

Although other approaches exist which differentiate between partial risk problems and risk or uncertainty problems, their major emphasis has not focussed on the unknown events themselves. Rather, a form of correction factor is assumed to apply after the expected utility criterion has been calculated according to probability laws. In all other approaches therefore the existence of probabilities for all events has been taken for granted.

The method used in Part II was to specify a set of axioms which a decision maker would be willing to follow if he accepts the expected utility criterion for decision problems under risk.

Most of the axioms are properties of the expected utility criterion. The axioms in Part II, therefore, reduce to the expected utility criterion if additional assumptions are made on the P-measure. However, many of the advantages of considering the more general set function would be lost.

References

Allais, M. "Le comportement de l'homme rationnel devant le risque: Critique des postulats et axioms de l'ecole Americaine", Econometrica 21:503-546 (1953).

Anscombe, F. and Aumann, R. J. "A definition of subjective probability", Ann. Math. Stat. 34:199-205 (1963).

Arrow, K. "Essays in the theory of risk-bearing", Markham Publishing Company, Chicago (1971).

Ash, R. B. "Real analysis and probability", Academic Press Inc., New York (1972).

Aumann, R. J. "Utility theory without the completeness axiom", Econometrica 30:445-461 (1962).

Becker, S. and Brownson, F. "What price ambiguity" or The role or ambiguity in decision-making", J. Polit. Econ. 72: 62-73 (1964).

Bernoulli, D. "Speciman Theoriae Novae de Mensura Sortis", Comentarii Acadeiae Scientiarum Imperiales Metropolitanae 5:175-192 (1738) (trans. by L. Sommer in Econometrica 22: 22-36 (1954).

Blum, J. R. and Rosenblatt, J. "On partial a priori information in Statistical inference", Ann. Math. Stat. 38:1671-1678 (1968).

Chipman, J. "The foundation of utility", Econometrica 28: 193-224 (1960).

Debreu, G. "Representation of a preference ordering by a numerical function", in R. M. Thrall, C. H. Coombs, and R. L. Davis (eds.) Decision Processes, John Wiley and Sons, New York, 159-165 (1954).

Debreu, G. "Stochastic choice and cardinal utility", Econometrica 26:440-444 (1958).

De Finetti, B. "Foresight: Its logical laws, its subjective sources" (trans by Kyburg and Smokler (eds.) in Studies in Subjective Probability (1964) 83-158, John Wiley and Sons, Inc., New York (1937).

DeGroot, M. H. "Optimal statistical decisions", McGraw-Hill, Inc., New York (1970).

Ellsberg, D. "Risk, ambiguity and the Savage axioms", Quart. J. Econ. 75:643-669 (1961).

Fellner, W. "Distortion of subjective probabilities as a reaction to uncertainty", Quart. J. Econ. 75:670-689 (1961).

Fishburn, P. "Analysis of decisions with incomplete knowledge of probabilities", Op. Res. 13:217-237 (1964).

Fishburn, P. "Utility theory for decision making", John Wiley and Sons, Inc., New York (1970).

Gardner, M. "Mathematical games", Scientific American 231: 102-109 (1974).

Good, I. J. "The estimation of probabilities: An essay on modern bayesian methods", Research Monograph No. 30, The MIT Press, Cambridge, Mass (1965).

Hagen, O. "New Foundations of utility: Allais versus Morgenstern", Working paper, Die Norweigsche Handelschochüle, Bergen, Norway, 15 pp. (1971)

Hagen, O. "A new axiomatization of utility under risk", Teorie a Metoda 4:55-80 (1972).

Halmos, P. R. "Measure theory", D. Van Nostran Company, Inc., Princeton, M. J. (1950).

Herstein, I. and Milnor, J. "An axiomatic approach to measurable utility", Econometrica 21:291-297 (1953).

Hurwitz, L. "Some specification problems and application to econometric models", Econometrica 19:343-344 (1951).

Jeffrey, R. "The logic of decision", McGraw-Hill, New York (1965).

Jensen, N. "An introduction to Bernoullian utility theory", Swedish Journal of Economics 69:163-183 (1967).

Knight, F. "Risk, uncertainty and profit", Houghton, Mifflin and Co., Boston (1921).

Kolmogorov, A. "Grundbegriffe der Wahrscheinlichkeitsrechnung", Berlin (1933).

Kraft, C., Pratt, J., and Seidenberg, A. "Intuitive probability on finite sets", Ann. Math. Stat. 30:408-419 (1959).

Lebesque, H. "Sur une généralisation de l'intégrale definie", Comp. Rend. 132:1025-1028 (1901).

Los, J. and Marczewski, E. "Extensions of measures", Fund. Math. 36:267-276 (1949).

Luce, R. and Krantz, D. "Conditional expected utility", Econometrica 39:253-271 (1971).

Luce, R. and Raiffa, H. "Games and decisions", John Wiley and Sons, Inc., New York (1957).

MacCrimmon, K. "An experimental study of the decision making behavior of business executives", unpublished dissertation, University of California, Los Angeles (1965).

MacCrimmon, K. and Larsson, S. "Utility theory: Axioms versus 'paradoxes'", Working Paper No. 311, University of British

Columbia (1975).

Marschak, J. "Rational behavior, uncertain prospects, and measurable utility", Econometrica 18:111-141 (1950).

May, K. "Intransitivity, utility, and the aggregation of preference patterns", Econometrica 22:1-13 (1954).

Menger, K. "Das Unsicherheitsmoment in der Wertlehre", Zeitschrift fur National-Okonie 4:459-485 (1950).

Menges, G. "On the 'bayesification' of the minimax principle", Unternehmenforschung 10:81-91 (1966).

Milnor, J. "Game against nature", in R. M. Thrall, C. H. Coombs, and R. L. Davis (eds.), Decision Processes, New York and London (1954).

Moskowitz, H. "Effects of problem representation and feedback on rational behavior in Allais and Morlat-type problems", Decision Sciences 5:225-241 (1974).

Nozick, R. "Newcomb's problem and two principles of choice", in Essays in Honor of Carl S. Hempel (ed. by N. Rescher) Dordrecht, Holland: Reidel Publishing Co., 114-146 (1969).

Peleg, B. "Utility functions for partially ordered topological spaces", Econometrica 38:93-96 (1970).

Pratt, J. W. "Risk aversion in the small and in the large", Econometrica 32:122-136 (1964).

Raiffa, H. "Decision analysis", Addison-Wesley Publishing Co., Reading, Mass. (1968).

Ramsey, F. P. "Truth and probability", in The foundations of mathematics and other logical essays (by R. P. Ramsey) Harcourt, Brace and Co., New York (1931), Reprinted in

Kyburg and Smokler (1964).

Randles, R. H. and HOllander, H. "γ-minimax selection procedures in treatments versus control problems", Ann. Math. Stat. 42:330-341 (1971).

Royden, H. L. "Real analysis", 2nd ed., MacMillan, New York (1968).

Samuelson, P. "Probability, utility and the independent axiom", Econometrica 29:670-678 (1952).

Savage, L. "The foundations of statistics", John Wiley and Sons, Inc., New York (1954).

Shuford, E. H. "A comparison of subjective probabilities for elementary and compound events", The Psychometric Laboratory, University of North Carolina, Chapel Hill, N. C., Number 20, January (1959).

Slovic, P. and Tversky, A. "Who accepts Savage's axiom?" Behavioral Science 19:368-373 (1975).

Tesfatsion, L. "Stochastic dominance and the maximization of expected utility", Discussion Paper No. 74-37, Center for Economic Research, Dept. of Economics, University of Minnesota (January, 1974).

Villegas, C. "On qualitative probability σ-algebras", Ann. Math. Stat. 35:1787-1796 (1964).

vonNeumann, J. and Morgenstern, D. "Theory of games and economic behavior, 2nd ed.", Princeton, N.J., Princeton University Press (1947).

Wald, A. "Statistical decision functions", John Wiley and Sons, Inc., New York (1950).

Yates, J. R. and Zukowski, L. G. "The anatomy and consequences of ambiguity in decision making", Michigan Mathematical Psychology Program:MMPP 1975-2 (1975).

Zadeh, L. A. "Fuzzy sets", Information and Control 8:338-353 (1965).

Zadeh, L. A. "Probability measures of fuzzy events", J. Math. Anal. and Appl. 23:421-427 (1968).

Zadeh, L. A. "Toward a theory of fuzzy systems", ERL Report No. 69-2, Electronics Research Laboratories, University of California, Berkeley, June (1969).

Appendix I

Some of the main theorems which guarantee the existence of the evaluation function f are summarized in this appendix. As was stated in section 3, the conditions for the existence of f are of two types -- one set of conditions on the order relation and the second on the topology on A. Because of their importance we shall give a brief summary of the types of relations to be considered.

Definition. Let S be a set. A binary relation Y on S is a subset of the Cartesian product S x S. Symbolically, $Y \subset S \times S$.

By this definition the empty set $\emptyset$ is a relation. This is the "smallest" relation since it is a subset of all other relations. Similarly, we have the "largest" relation L = SxS which contains all relations on S x S. Other relations which occur often enough to deserve mention are the diagonal relation, ∇, of a set S, defined by $\nabla = \{(x,y) \in SxS | x = y\}$, and the dual relation Y' of a relation Y, defined by $Y' = \{(x,y) | (y,x) \in Y\}$. The dual is, therefore, the mirror image across the diagonal. The complement relation, $\overline{Y}$ of a relation Y, is the complement of the set Y, i.e., $\overline{Y} = L-Y$, the usual complement used in set theory.

Multiplication of two relations can also be defined. Let C and D be two relations on the same set S. The product denoted by $C \bullet D$, is defined by

$$C \bullet D = \{(x,y) | (x,z) \in C, (z,y) \in D, \text{ for some } z \in S\}.$$

We shall not discuss the properties of this product, but we shall use it in some definitions. Some relations have properties occurring so often that names have been given to these properties. We shall summarize these here and use the definitions by Chipman (1960) for the more common ones.

Definition. Letting M be any relation on a set S we define the terms:

1. Transitive: $M \cdot M \subset M$
2. Negative Transitive: $\overline{M} \cdot \overline{M} \subset \overline{M}$
3. Reflexive: $\nabla \subset M$
4. Irreflexive: $\nabla \subset \overline{M}$
5. Comparability: $L \subset M \cup M'$
6. Symmetric: $M' \subset M$
7. Asymmetric: $M' \subset \overline{M}$
8. Antisymmetric: $M \cap M' \subset \nabla$

Definition. A relation M is a partial ordering on a set S provided M is relfexive, antisymmetric and transitive.

It is common to write $x \overset{S}{\leqslant} y$ (or $y \overset{S}{\geqslant} x$) rather than $(x,y) \in M$. We shall adopt the same notation here. As a rule we shall denote any ordering by $\leqslant$ (or $>$) if it is obvious on which set the ordering is defined. Where doubt may occur we indicate the set along with the ordering, i.e., $\overset{A}{\leqslant}$ is an order relation on A.

Associated with any partial ordering $\overset{S}{\leqslant}$ we can define a relation $\overset{S}{<}$ by: $x \overset{S}{<} y$ if and only if $x \overset{S}{\leqslant} y$ and not $y \overset{S}{\leqslant} x$. Clearly $\overset{S}{<}$ is transitive and irreflexive, and is sometimes used as a definition of a partial ordering. Here we will call it a <u>strict partial order</u>. We can also define a relation $\overset{S}{=}$ by:

$$x \overset{S}{=} y \text{ if and only if } x \overset{S}{\leqslant} y \text{ and } y \overset{S}{\leqslant} x.$$

The second type of ordering we shall consider is a weak ordering.

<u>Definition</u>. A relation $\overset{S}{<}$ is a <u>weak ordering</u> provided $\overset{S}{<}$ is asymmetric and negative transitive.

Since an asymmetric and negative transitive binary relation is irreflexive and transitive a strict partial ordering is contained in a weak ordering. If M is either a strict partial ordering or a weak ordering, then we define $\overset{M}{=}$ by:

$x \overset{M}{=} y$ if not $x \overset{M}{<} y$ and not $y \overset{M}{<} x$ and define $\overset{M}{\leqslant}$ by:

$$x \overset{M}{\leqslant} y \text{ if } x \overset{M}{<} y \text{ or } x \overset{M}{=} y.$$

In doing so, we gain the comparability property, i.e., given any $x,y \in M$ then either $x \overset{M}{\leqslant} y$ or $y \overset{M}{\leqslant} x$. Thus, we have reached the ordering that real numbers satisfy except for perhaps the antisymmetry property.

<u>Definition</u>. A relation $\overset{M}{\leqslant}$ is a <u>complete ordering</u> if it is comparable and transitive.

It will, therefore, not come as a surprise if the ordering we need for the existence of f is either a weak ordering or a partial ordering.

Topological necessities. If we only assume that the ordering given is either a weak ordering or a strict partial ordering on a set A, then it is not sufficient to specify an order preserving real-valued function. We have to make additional assumptions on the cardinality of the set A.

Theorem 1. If $\overset{A}{<}$ is a weak ordering on A and the cardinality of A is countable then there is a real-valued function f on A such that

$$a \overset{A}{<} b \text{ implies } f(a) < f(b) \text{ for all } a,b \in A.$$

The result would still hold if we replaced the weak ordering by a strict partial ordering. We shall not prove this theorem here as the proof is easy and can be found in most books on utility theory. It was first proved in this form by Debreu (1954). It is not difficult however to see that the cardinality assumption in Theorem 1 could be weakened. We would only need to have the set of equivalence classes induced by the weak ordering to be countable. In this case, we induce an ordering on the equivalence classes by the ordering $\overset{A}{<}$.

If we now remove the condition that A must be countable (or that the set of equivalence classes must be countable), we must still keep some control over the cardinality of A. The condition required is the existence of a countable, dense subset of A. This was used by Fishburn (1970). Peleg (1970) called this condition a separability condition. Here, we shall use Fishburn's terminology since separability as used *in* topology means something quite different.

<u>Definition</u>. Let $\overset{A}{\leqslant}$ be a binary relation on A, and let Z be a subset of A. Then Z is <u>order-dense</u> in A if and only if whenever $a \overset{A}{\leqslant} b$ there is a $z \in Z$ such that $a \overset{A}{\leqslant} z$ and $z \overset{A}{\leqslant} b$.

Let A be the set of all real numbers and Z the set of all rational numbers, and let the ordering defined by the natural ordering of numbers on both A and Z. Then Z is order-dense since between every two distinct numbers there exists a rational number.

This condition controls the cardinality of the set A if the cardinality of Z is specified and is sufficient for proving the second theorem (Fishburn, 1970, p.27).

<u>Theorem 2</u>. If A is weakly ordered and there is a countable subset $Z \subset A$ which is order-dense in A, then there is a real-valued function f on A such that

$$a \overset{A}{<} b \text{ implies } f(a) < f(b) \text{ for all } a,b \in A.$$

We would now like to relate the concept of order-dense sets to more intuitive concepts in topology. To do so we must introduce some of the ideas in topology. One of the fundamental concepts of topology is that of "nearness". One natural way of defining what we mean by "action a is near action b" for some $a,b \in A$, is by inducing a distance measure on A. Let us first assume that there exists a real-valued function f on A. A weak topology may then be defined on A, be assuming that for any open set of real numbers N containing $f(a)$, the set $\{b \varepsilon A | f(b) \varepsilon N\}$ is defined to be open for all $a \varepsilon A$.

In doing so, f would always be continuous and the concept of "nearness" is inherited from "nearness" on the real line. This definition is equivalent to saying that the sets $\{z \varepsilon A | f(z) < f(a)\}$ and $\{z \varepsilon A | f(a) < f(z)\}$ are open for all $a \varepsilon A$. It also implies that if $a_i, i=1,2,...$ is a sequence of actions, $\{a_i\}$ converges to a if and only if

$$\lim_{i \to \infty} f(a_i) = f(a).$$

If the function f is not given, but a weak ordering exists on A such that $a,b \varepsilon A$ implies either $a \overset{A}{<} b$ or $b \overset{A}{<} a$, then a topology T may be defined on A by letting the sets $\{z \varepsilon A | a \overset{A}{<} z\}$and $\{z \varepsilon A | z \overset{A}{<} a\}$ for all $a \varepsilon A$ form a subbase for the topology T. We shall denote the closure of the set $\{z \varepsilon A | a \overset{A}{<} z\}$ by $\{z \varepsilon A | a \overset{A}{\leqslant} z\}$. We now show how this can be related back to order dense sets. The condition necessary for this is: if A is separable and connected (in a topological sense), then there exists a countable subset Z of A which is order dense in A.

To see this, let $a \overset{A}{<} b$. Then the sets $\{z \varepsilon A | z \overset{A}{\leqslant} a\}$ and $\{z \varepsilon A | b \overset{A}{\leqslant} z\}$ are disjoint, closed and non-empty and therefore:

$$\{z \varepsilon A | z \overset{A}{<} b\} \cap \{z \varepsilon A | a \overset{A}{<} z\}$$

is open and non-empty. Since A is separable it must contain a countable dense subset Z such that there exists a $z \in Z$ satisfying $a \overset{A}{<} z$ and $z \overset{A}{<} b$.

Therefore if A is weakly ordered, and is separable and connected, this implies that there exists a countable subset $Z \subset A$ which is order dense in A. Also note that the existence of such a set Z is all that is needed for the existence of a real-valued function f on A such that

$$a \overset{A}{<} b \text{ implies } f(a) < f(b) \text{ for all } a,b \in A$$

by Theorem 2. This gives us the following theorem (Debreu, 1954).

> <u>Theorem 3</u>. If A is weakly-ordered, separable and connected, with the natural topology T, then there exists a real-valued function f on A, such that: if $a \overset{A}{<} b$ then $f(a) < f(b)$ for all $a,b \in A$.

Similarly, Peleg (1970) has proved the same theorem for strict partial ordering:

> <u>Theorem 4</u>. If A is a strict partial ordering on a connected, separable set A with the natural topology T, then there exists a real-valued function f on A such that: if $a \overset{A}{<} b$ then $f(a) < f(b)$ for all $a,b \in A$.

A summary of conditions necessary for the existence of f now follows:

1) Ordering assumption, i.e., transitivity, partial, weak

or complete.

2) Topology assumption, i.e., the cardinality assumption on A.

From a decision maker's point of view we would therefore be most interested in the transitivity and completeness, since the topological assumption does not affect preferences among alternatives.

Appendix II

The purpose of this appendix is to summarize some of the basic definitions and results concerning measure and function theory.

1. Measures and the extension theorem

The class Θ of subsets of Ω which has the following properties:

a) $\Omega \varepsilon \Theta$

b) if $B \varepsilon \Theta$ then $\bar{B} \varepsilon \Theta$, where $\bar{B} = \Omega - B$

c) if $B_1, B_2, \ldots$ then $\bigcup_{i=1}^{\infty} B_i \varepsilon \Theta$

is called a <u>σ-field</u> or <u>σ-algebra</u>.

A <u>measurable space</u> is a set Ω, and a σ-algebra Θ of subsets of Ω. A <u>measure</u> on a σ-field Θ is a non-negative, extended real-valued function μ such that whenever $B_1, B_2, \ldots$ form a finite or countably infinite collection of disjoint sets in Θ we have

$$\mu\left(\bigcup_n B_n\right) = \sum_n \mu(B_n).$$

If $\mu(\Omega) = 1$, μ is called a probability measure.

A <u>measure space</u> is a triple (Ω, Θ, μ) where Ω is a set, Θ is a σ-algebra of subsets of Ω, and μ is a measure on Θ. If μ is a probability measure (Ω, Θ, μ) is called a probability space.

The definition of a probability measure implies the following consequences where all sets are members of Θ.

1. $\mu(\phi) = 0$.

2. $\mu(\bar{E}) = 1 - \mu(E)$.

3. $\mu(E \cup F) + \mu(E \cap F) = \mu(E) + \mu(F)$.

4. $E \subset F$ implies that $\mu(E) = \mu(F) - (F-E) \leqslant \mu(F)$.

5. Monotone property. If E_n is an increasing sequence of set or a decreasing sequence of set converging to E, then $\mu(E_n) \to \mu(E)$.

6. Boole's inequality. $\mu(\bigcup_i E_i) \leqslant \sum_i \mu(E_i)$.

On each subset D of Ω, we can define an outer measure by

$$\mu^*(D) = \inf_{M \supset D} \mu(M) \qquad M \in \Theta$$

and an inner measure by

$$\mu_*(D) = \sup_{M \subset D} \mu(M) \qquad M \in \Theta$$

The following properties of the inner and outer measure are easily verified (or can be found in Halmos, 1950, Lebesque, 1901, or Royden, 1968).

i) $\mu_*(D) \leqslant \mu^*(D)$ for all $D \in \Omega$

ii) if $C \cup D \in \Theta$, $C \cap D = \emptyset$ then $\mu_*(C) + \mu^*(D) = \mu(C \cup D)$

iii) if $C \cap D = \emptyset$, then $\mu_*(C \cup D) \leqslant \mu_*(C) + \mu^*(D) \leqslant$
$\leqslant \mu^*(C \cup D)$

iv) $\mu_*(D) = \mu^*(D) = \mu(D)$ if $D \in \Theta$.

For $D \in \Theta$ and $D \cap C = \emptyset$ the following six properties hold:

v) $\mu^*(D \cup C) = \mu^*(D) + \mu^*(C)$

vi) $\mu_*(D \cup C) = \mu_*(D) + \mu_*(C)$

vii) $\mu^*(C) = \mu^*(C \cup D) + \mu^*(C \cup \bar{D})$

viii) $\mu_*(C) = \mu_*(C \cup D) + \mu_*(C \cup \bar{D})$

ix) $\mu^*(C) = 1 - \mu_*(\bar{C})$

x) if $D \subset F$ and $C \subset E$ where $F \cap E = \emptyset$ and $F, E \in \Theta$ then

$$\mu^*(C \cup D) = \mu^*(C) + \mu^*(D)$$

$$\mu_*(C \cup D) = \mu_*(C) + \mu_*(D).$$

A more general definition of an outer measure μ^* is sometimes used. It is an extended real-valued set function defined on all subsets of a space Ω and having the following properties:

i) $\mu^*(\emptyset) = 0$

ii) if $A \subset B$ then $\mu^*(A) \leqslant \mu^*(B)$

iii) if $E \subset \bigcup_{i=1}^{\infty} E_i$ then $\mu^*(E) \leqslant \sum_{i=1}^{\infty} \mu^*(E_i)$.

The second property is called monotonicity and the third countable subadditivity. In view of (i) finite subadditivity follows from (iii).

From an outer measure a new measure μ_1 can be defined. In this case the difficulty arises in making sure from which sets in Ω the additivity properties hold. This is usually done by specifying a class of measurable sets in the following way: A set E is <u>measurable</u> with respect to μ^* if for every set A we have

$$\mu^*(A) = \mu^*(A \cap E) + \mu^*(A \cap \bar{E}).$$

This guarantees that μ^* is additive out of this class of sets, and also that the class of measurable sets is a σ-algebra (Royden, 1968, Theorem 1, p.251).

<u>Theorem AI</u>: The class Ψ of μ^*-measurable sets is a σ-algebra.

If we start with an outer measure μ^*, and define a measure μ_1 from the μ^*, and then induce an outer measure $\overline{\mu}_1$ from μ_1, and if it so happens that $\mu^* = \overline{\mu}_1$ we say μ^* is a <u>regular outer measure</u>.

A regular outer measure μ^* has the property that if B_i is a sequence of increasing sets such that $\bigcup B_i = B$ then

$$\lim_i \mu^*(B_i) = \mu^*(B).$$

This is the property needed in Axiom V (Part II of the thesis).

2. Measurable functions

Given two spaces Ω, R, and a mapping $X(\omega):\Omega \to R$, the inverse image of a set $B \subset R$ is defined as

$$X^{-1}B = \{\omega \in \Omega; X(\omega) \in B\}.$$

Denote this by $\{X \in B\}$. The taking of inverse images preserves all set operations; that is

$$\{X \in \bigcup_\lambda B_\lambda\} = \bigcup_\lambda \{X \in B_\lambda\},$$

$$\{X \in \bigcap_\lambda B_\lambda\} = \bigcap_\lambda \{X \in B_\lambda\},$$

$$\{X \in \bar{B}\} = \overline{\{X \in B\}}.$$

Given two measurable spaces (Ω,Θ), (R,Ψ), a function $X:\Omega \to R$ is called measurable if the inverse of every set in Ψ is in Θ. Therefore if a measure is defined on Θ, an <u>induced measure</u> P may be defined on Ψ by the measurable mapping X as follows:

$$P(B) \equiv \mu(X \in B) \quad \text{for all } B \in \Psi.$$

By the properties of the inverse image P is guaranteed to be a measure.

If a sequence of measurable functions converges to a function, then this function is also measurable (Royden, 1968, Theorem 6, p.223).

<u>Theorem AII</u>. The class of Θ-measurable functions is closed under pointwise convergence. That is, if $X_n(\cdot)$ are Θ-measurable for each n, and $\lim_n X_n(\omega)$ exists for every ω, then $X(\omega) = \lim_n X_n(\omega)$ is Θ-measurable.

3. Integrations

Let $X(\cdot)$ be a function from Ω to R, then $X(\cdot)$ is a <u>simple function</u> if there exists a finite number of subsets of Ω, denoted by $B_1,\ldots,B_n$ such that

$$B_i \cap B_j = \emptyset \text{ for } i\neq j, \ \Omega \subset \bigcup_i B_i$$

$$\text{and } X_{B_i}(\cdot) = r_i \quad \text{for all } i=1,\ldots,n.$$

To define the integral we first define the integral for simple functions, then non-negative functions and lastly for arbitrary functions.

The <u>integral</u>. Take (Ω,Θ,μ) to be a measure space, and let R be the real line, then $\int X d\mu$ of the non-negative measurable simple function is defined by $\sum_{i=1}^{n} r_i \mu(B_i)$. Let $X(\omega) \geqslant 0$ be a non-negative Θ-measurable function. To define the integral of X let $X_n \geqslant 0$ be simple functions such that X_n increases to X.

For X_n increasing to X, it is easy to show that $\int X_{n+1} d\mu \geqslant \int X_n d\mu > 0$.

Define $\int X d\mu$ as $\lim_n \int X_n d\mu$. Furthermore, the value of this limit is the same for all sequences of non-negative simple functions converging up to X (Halmos, p.101).

If $\int |X| d\mu < \infty$, define

$$\int X d\mu = \int X^{+} d\mu - \int X^{-} d\mu;$$

where $X(\cdot)^{+} = \max\{X(\cdot,0)\}$ and $X(\cdot)^{-} = \max\{-X(\cdot),0\}$.
The elementary properties of the integral are: If the integrals of X and Y exist,

i) $X \geqslant Y$ implies $\int X d\mu \geqslant \int Y d\mu$,

ii) $\int (\alpha X + \beta Y) d\mu = \alpha \int X d\mu + \beta \int Y d\mu$,

iii) $A,B \in \Theta, A \cap B = \emptyset$ implies $\int_{A \cup B} X d\mu = \int_{A} X d\mu + \int_{B} X d\mu$.

Some nonelementary properties are:

<u>Monotone Convergence Theorem AIII</u>. (Halmos, 1950, Theorem B, p.112). For $X_n > 0$ non-negative increasing

Θ-measurable functions, which converge to X, then

$$\lim \int X_n d\mu = \int X d\mu .$$

Theorem AIV. (Halmos, 1950, Theorem I, p.98). Let $X(\cdot)$ be a Borel measurable function such that $\int_\Omega X d\mu$ exists. If B_i is a set of disjoint subsets of Ω then

$$\int_\Omega X d\mu = \sum_i \int_{B_i} X d\mu .$$

This summarizes the basic definitions and results needed in Part I and Part II of the thesis.

Appendix III

In this appendix we shall summarize some of the different approaches to the expected utility theory. Before stating the axiom we shall summarize the notations used concerning the reward functions.

For each $a \varepsilon A$, a function $X(\cdot,a)$ from Ω to R is called a reward function.

If $Y(\cdot)$ is a function from B to R where $B \subset \Omega$, then Y is called a restriction of $X(\cdot,a)$ if $X(\cdot,a) = Y(\cdot)$ for all $\omega \varepsilon B$. The function $Y(\cdot)$ will, in this case, be denoted by $X_B(\cdot,a)$.

Let $X_B(\cdot,a)$ and $X_C(\cdot,c)$ be two restrictions of the reward functions $X(\cdot,a)$ and $X(\cdot,c)$ if there exists a function $Y(\cdot)$ from $B \cup C$ to R such that $Y(\cdot) = X_B(\cdot,a)$ and $Y(\cdot) = X_C(\cdot,c)$. It will be convenient at some cases to denote the function $Y(\cdot)$ by $X_{B \cup C}(\cdot,ac)$, especially so in the Luce and Krantz approach.

<u>Approaches to expected utility theory</u>.

The different approaches we have considered were developed by:

1) vonNeumann-Morganstern (1947)
2) Marschak (1959)
3) Savage (1954)
4) Arrow (1971)
5) Luce and Krantz (1971)

We shall summarize the axioms assumed by each of these.

vonNeumann-Morgenstern Axioms. The vonNeumann-Morgenstern (1947) approach does not directly make any assumptions on the underlying probability space (Ω,Θ,μ) or on Γ. It does assume that Π, the set of all induced probability measures on R, is equal to the set of all discrete probability measures on R. They induce an ordering on Π from an ordering on R. To present their approach, we need some new definitions. Let $\Pi'' = (0,1) \times R \times R$, and let F be a function from Π'' to R. The set Π'' can be thought of as the set of those induced probability measures on R which are non-zero for exactly two elements of R. The value $F(\alpha,r_1,r_2) = r_3$ can be thought of as the reward $r_3 \in R$ which would make us indifferent between the gambles.

1. receiving r_1 with the probability α
   receiving r_2 with the probability $1-\alpha$
2. receiving r_3 with the probability 1

In what follows, it is assumed that $r_1,r_2,r_3 \in R$ and that α, β and γ are real numbers on (0,1).

Axiom NM1: $\overset{R}{<}$ is a complete, hereditary ordering.

Axiom NM2: $r_1 \overset{R}{<} r_2$ implies $r_1 \overset{R}{<} F(\alpha r_1,r_2)$ and

$$F(\alpha,r_1,r_2) \overset{R}{<} r_2 \text{ for all } \alpha \in (0,1)$$

Axiom NM3: $r_1 \overset{R}{<} r_3 \overset{R}{<} r_2$ implies the existence of an $\alpha \in (0,1)$ and a $\gamma \in (0,1)$ with $F(\alpha,r_1,r_2) \overset{R}{<} r_3$ and

$r_3 \overset{R}{<} F(\gamma, r_1, r_2)$.

Axiom NM4: $F(\alpha, r_1, r_2) \overset{R}{=} F(1-\alpha, r_2, r_1)$ and $F(\alpha, F(\gamma, r_1, r_2), r_2) \overset{R}{=} F(\alpha\gamma, r_1, r_2)$ for all $\alpha, \gamma \in (0,1)$.

These assumptions are sufficient to prove that a real valued function U exists such that

$$r_1 \overset{R}{<} r_2 \text{ implies } U(r_1) < U(r_2)$$

and

$$U(R\alpha, r_1, r_2)) = \alpha U(r_1) + (1-\alpha) U(r_2).$$

Marschak Axioms. Marschak (1950) was the first to adopt an approach of establishing an ordering on the probability measures. Samuelson (1952), Herstein and Milnor (1953) and other authors have also adopted this formulation. The axioms we shall give here are essentially the same as Jensen's (1964) axioms, and he has shown them to imply Marschak's axioms.

In this approach also, we ignore the underlying probability space since all assumptions are based on the induced probability measures. R is assumed to be a finite set. We denote Π, as before, to be the set of all probability measures on R.

Axiom M1: $\overset{\Pi}{<}$ is a complete, hereditary ordering.

Axiom M2: If $P_1, P_2, P_3 \in \Pi$ and $P_1 \overset{\Pi}{<} P_2$, then for any real number $\alpha \in (0,1)$,

$$\alpha P_1 + (1-\alpha)P_3 \overset{\Pi}{<} \alpha P_2 + (1-\alpha)P_3$$

Axiom M3: If $P_1 \overset{\Pi}{<} P_2 \overset{\Pi}{<} P_3$ then there exists real numbers $\alpha, \beta \in (0,1)$ such that $\alpha P_1 + (1-\alpha)P_3 \overset{\Pi}{<} P_2$ and $P_2 \overset{\Pi}{<} \beta P_1 + (1-\beta)P_3$.

These three assumptions are sufficient to derive a utility function.

Savage Axioms. Savage (1954) starts with an underlying probability space (Ω, Θ, μ). The theory does not hold, however, for an arbitrary probability space. There are restrictions on the cardinality of A (Axiom S6) and Θ must contain all subsets of Ω. The probability μ is not explicitly defined but is derived from preference relations on subsets of Ω.

Axiom S1: $\overset{\Gamma}{<}$ is a complete hereditary ordering.

Let $B, \overline{B} \in \beta$ and $r_1, r_2 \in R$ and $a,b,c,d \in A$.

Axiom S2: If $X_B(\cdot,a) = X_B(\cdot,b)$, $X_B(\cdot,c) = X_B(\cdot,d)$,

$$X_B(\cdot,a) = X_{\overline{B}}(\cdot,c),\ X_{\overline{B}}(\cdot,b) = X_{\overline{B}}(\cdot,d)$$

$$\text{and } X(\cdot,a) \overset{\Gamma}{\leqslant} X(\cdot,b)$$

$$\text{then } X(\cdot,c) \overset{\Gamma}{\leqslant} X(\cdot,d).$$

Axiom S3: Let $X(\cdot,a) = r_1$ and $X(\cdot,b) = r_2$

$$\text{If } X_B(\cdot,c) = X_B(\cdot,a),\ X_B(\cdot,d) = X_B(\cdot,b)$$

$$\text{and } X_{\overline{B}}(\cdot,c) = X_{\overline{B}}(\cdot,d)$$

$$\text{then } X(\cdot,a) \overset{\Gamma}{\leqslant} X(\cdot,b) \text{ if and only if}$$

$X(\cdot,c) \overset{\Gamma}{\leqslant} X(\cdot,d)$ for all B such that B is not null[2].

<u>Axiom S4</u>: If $B,C \subset \Omega$ and $r_i \in R$ $i=1,2,3,4$, $r_1 > r_2$, $r_3 > r_4$ $a,b,c,d \in A$, and

$X_B(\cdot,a) = r_1$	$X_{\overline{B}}(\cdot,a) = r_2$
$X_C(\cdot,b) = r_1$	$X_{\overline{C}}(\cdot,b) = r_2$
$X_B(\cdot,c) = r_3$	$X_{\overline{B}}(\cdot,c) = r_4$
$X_C(\cdot,d) = r_3$	$X_{\overline{C}}(\cdot,d) = r_4$

$$\text{and } X(\cdot,a) \overset{\Gamma}{\leqslant} X(\cdot,b)$$

$$\text{then } X(\cdot,c) \overset{\Gamma}{\leqslant} X(\cdot,d).$$

<u>Axiom S5</u>: There is at least one pair of rewards $r_1, r_2 \in R$ such that if $a,b \in A$ are defined by $X(\cdot,a) \equiv r_1$ and $X(\cdot,b) \equiv r_2$, then

$$X(\cdot,a) = r_1 \overset{\Gamma}{<} X(\cdot,b) = r_2.$$

<u>Axiom S6</u>: If $X(\cdot,a) \overset{\Gamma}{<} X(\cdot,b)$ and r is any reward in R, then there exists a partition P of Ω, such that for any $B \in P$

---

[2] A set $B \in \Theta$ is null if $\mu(B) = 0$, or in Savage terminology, B is null if $X_B(\cdot,a) \overset{\Gamma}{=} X_B(\cdot,c)$ for all $a,c \in A$.

$$X_B(\cdot,c) = r,\ X_{\overline{B}}(\cdot,c) = X_{\overline{B}}(\cdot,a) \text{ implies}$$

$$X(\cdot,c) \overset{\Gamma}{<} X(\cdot,b) \text{ and}$$

$$X_B(\cdot,d) = r,\ X_{\overline{B}}(\cdot,d) = X_{\overline{B}}(\cdot,b) \text{ implies}$$

$$X(\cdot,a) \overset{\Gamma}{<} X(\cdot,d).$$

Axiom S7: If $X_B(\omega,a) \overset{\Gamma}{<} X(\omega,b)$ for all $\omega \in B$, then

$$X_B(\cdot,a) \overset{\Gamma}{<} X_B(\cdot,b).$$

The first six axioms are sufficient to guarantee a probability μ and a function U such that the expected utility preserves the ordering when the reward set is finite. To extend the result to an infinite reward set, Axiom S7 is also needed.

Arrow's Axioms. Arrow (1971) does not make the restrictions on the probability space (Ω,Θ,μ) that Savage does but his overall approach is similar to Savage's.

Axiom A1: $\overset{A}{<}$ is a complete, hereditary ordering.

Axiom A2: Given $a,b \in A$ where $b \overset{A}{<} a$, a reward $r \in R$, and $\{E^i\}$ a sequence of sets in Θ such that $E^{i+1} \subset E^i$ with $\bigcap E_i = \emptyset$. Define actions $a^i \in A$, $b^i \in A$ by $X_{\overline{E}i}(\cdot,a^i) = X_{\overline{E}i}(\cdot,a)$, $X_{Ei}(\cdot,a^i) = c$.

$$X_{\overline{E}i}(\cdot,b^i) = X_{\overline{E}i}(\cdot,b),\ X_{Ei}(\cdot,b^i) = c.$$

Then for all i sufficinetly large

$$b \overset{A}{<} a^i \text{ and } b^i \overset{A}{<} a.$$

<u>Axiom A3:</u> For any given event E, $\overset{A}{<}$ satisfies Axiom A2 such that any two actions $a,b \in A$ where $X_E(\cdot,a) = X_E(\cdot,b)$ will be indifferent given E.

This is denoted as $a \overset{A}{=} b|E$. Weak preference is denoted as $a \overset{A}{<} b|E$ and strict preference as $a \overset{A}{<} b|E$.

<u>Axiom A4</u>: Let P be a partition. Given two actions $a,b \in A$, if for every $E \in P$, $b \overset{A}{<} a|E$, then $b \overset{A}{<} a$. If in addition, there is a collection P' of events in P, whose union is non-null, such that $b \overset{A}{<} a|E$, $E \in P'$, then $b \overset{A}{<} a$.

<u>Axiom A5</u>: If $a,b,c,d \in A$ and

$$X(\omega_1,a) = X(\omega_2,b),\ X(\omega_1,c) = X(\omega_2,d),$$

$$c \overset{A}{<} a|\omega_1 \text{ implies } d \overset{A}{<} b|\omega_2.$$

<u>Axiom A6</u>: The probability distribution of states of the world is atomless. If the probability distribution of consequences is the same for two actions, they are indifferent.

This set of axioms is also sufficient to prove the existence of a real-valued utility function U on R such that the ordering on A satisfies the expected utility criterion.

Luce and Krantz Axioms. One argument criticizing Savage's axioms has been that all the random variables have been defined on the same state space. For example, in betting that heads will occur when a coin is flipped, or when considering investing in a particular stock, clearly the state of the world is quite different. Luce and Krantz (1971) developed an axiom system to handle this case, as follows: Let Φ denote a set of functions from non-null events and Θ a given algebra and ¶ the null events in Θ. Using this notation their axioms can be stated as:

Assume that $C,B \in \Theta$.

Axiom LK1: i) If $D \cap B = \emptyset$ then $X_{D \cup B}(\cdot, ac) \in \Phi$

ii) If $B \subset D$, $X_D(\cdot, a) \in \Phi$ implies

$$X_B(\cdot, a) \in \Phi.$$

Axiom LK2: $\overset{\Phi}{\geqslant}$ is a weak ordering.

Axiom LK3: If $D \cap B = \emptyset$ and $X_D(\cdot, a) \overset{\Phi}{=} X_B(\cdot, b)$ then

$$X_{D \cup B}(\cdot, ab) \overset{\Phi}{=} X_D(\cdot, a).$$

Axiom LK4: If $D \cap B = \emptyset$ then

$$X_D(\cdot, a) \overset{\Phi}{\geqslant} X_D(\cdot, b) \text{ if and only if}$$

$$X_{D \cup B}(\cdot, ac) \overset{\Phi}{\geqslant} X_{D \cup B}(\cdot, bc).$$

Axiom LK5: If $D \cap B = \emptyset$, $X_D(\cdot,i) \overset{\Phi}{=} X_B(\cdot,i)$ $i=a,b,c,d$

$$X_{D\cup B}(\cdot,ag) \overset{\Phi}{=} X_{D\cup B}(\cdot,bh) \text{ and}$$

$$X_{D\cup B}(\cdot,ka) \overset{\Phi}{=} X_{D\cup B}(\cdot,lb) \text{ then}$$

$$X_{D\cup B}(\cdot,cg) \overset{\Phi}{\geqslant} X_{D\cup B}(\cdot,dh) \text{ if and only if}$$

$$X_{D\cup B}(\cdot,kc) \overset{\Phi}{\geqslant} X_{D\cup B}(\cdot,ld).$$

Axiom LK6: If $D \cap B = \emptyset$, and for any sequence of action $i \in M$,

$$X_B(\cdot,a_i) \overset{\Phi}{\neq} X_N(\cdot,a_2) \text{ and}$$

$$X_{D\cup B}(\cdot,a_i a_1) \overset{\Phi}{=} X_{D\cup B}(\cdot,a_{i+1}a_2) \text{ for all } i,$$

then either M is finite or $\{X_D(\cdot,a_i) \mid i \in M\}$ is unbounded.

Axiom LK7: i) If $B \in ¶$ and $D \subset B$, then $D \in ¶$.

ii) $B \in ¶$ if and only if for all

$$X_{D\cup B}(\cdot,ab) \in \Phi,\ X_{D\cup B}(\cdot,ab) \overset{\Phi}{=} X_D(\cdot,a).$$

Axiom LK8: $\Theta - ¶$ contains at least three pairwise disjoint elements. Φ contains at least two actions such that $X_D(\cdot,a) \overset{\Phi}{\neq} X_B(\cdot,b)$.

<u>Axiom LK9</u>: i) If C and $X_B(\cdot,a)$ are given then there exists $X_C(\cdot,d) \in \Phi$ for which $X_B(\cdot,a) \overset{\Phi}{=} X_C(\cdot,d)$.

ii) If $D \cap B = \emptyset$ and $X_{D\cup B}(\cdot,ac) \geqslant X_{D\cup B}(\cdot,d) \geqslant X_{D\cup B}(\cdot,bc)$, then there exists $X_D(\cdot,e)$ such that $X_{D\cup B}(\cdot,d) \overset{\Phi}{=} X_{D\cup B}(\cdot,ec)$.

This completes the different axioms for the different approaches considered here.