

Decision Theory

A Brief Introduction

1994-08-19

Minor revisions 2005-08-23

Sven Ove Hansson

Department of Philosophy and the History of Technology

Royal Institute of Technology (KTH)

Stockholm

Contents

Preface	4
1. What is decision theory?	5
1.1 Theoretical questions about decisions	5
1.2 A truly interdisciplinary subject.....	6
1.3 Normative and descriptive theories.....	6
1.4 Outline of the following chapters.....	8
2. Decision processes.....	9
2.1 Condorcet	9
2.2 Modern sequential models	9
2.3 Non-sequential models.....	10
2.4 The phases of practical decisions – and of decision theory	12
3. Deciding and valuing.....	13
3.1 Relations and numbers	13
3.2 The comparative value terms	14
3.3 Completeness	16
3.4 Transitivity	17
3.5 Using preferences in decision-making.....	19
3.6 Numerical representation	20
3.7 Using utilities in decision-making	21
4. The standard representation of individual decisions.....	23
4.1 Alternatives	23
4.2 Outcomes and states of nature	24
4.3 Decision matrices	25
4.4 Information about states of nature	26
5. Expected utility.....	29
5.1 What is expected utility?.....	29
5.2 Objective and subjective utility.....	30
5.3 Appraisal of EU.....	31
5.4 Probability estimates	34
6. Bayesianism.....	37
6.1 What is Bayesianism?	37
6.2 Appraisal of Bayesianism	40
7. Variations of expected utility	45
7.1 Process utilities and regret theory	45

7.2 Prospect theory	47
8. Decision-making under uncertainty	50
8.1 Paradoxes of uncertainty	50
8.2 Measures of incompletely known probabilities	52
8.3 Decision criteria for uncertainty	55
9. Decision-making under ignorance.....	59
9.1 Decision rules for "classical ignorance"	59
9.2 Unknown possibilities.....	63
10. The demarcation of decisions.....	68
10.1 Unfinished list of alternatives	68
10.2 Indeterminate decision horizons	69
11. Decision instability.....	73
11.1 Conditionalized EU.....	73
11.2 Newcomb's paradox	74
11.3 Instability.....	76
12. Social decision theory.....	79
12.1 The basic insight	79
12.2 Arrow's theorem.....	81
References	82

Preface

This text is a non-technical overview of modern decision theory. It is intended for university students with no previous acquaintance with the subject, and was primarily written for the participants of a course on risk analysis at Uppsala University in 1994.

Some of the chapters are revised versions from a report written in 1990 for the Swedish National Board for Spent Nuclear Fuel.

Uppsala, August 1994
Sven Ove Hansson

1. What is decision theory?

Decision theory is theory about decisions. The subject is not a very unified one. To the contrary, there are many different ways to theorize about decisions, and therefore also many different research traditions. This text attempts to reflect some of the diversity of the subject. Its emphasis lies on the less (mathematically) technical aspects of decision theory.

1.1 Theoretical questions about decisions

The following are examples of decisions and of theoretical problems that they give rise to.

Shall I bring the umbrella today? – The decision depends on something which I do not know, namely whether it will rain or not.

I am looking for a house to buy. Shall I buy this one? – This house looks fine, but perhaps I will find a still better house for the same price if I go on searching. When shall I stop the search procedure?

Am I going to smoke the next cigarette? – One single cigarette is no problem, but if I make the same decision sufficiently many times it may kill me.

The court has to decide whether the defendant is guilty or not. – There are two mistakes that the court can make, namely to convict an innocent person and to acquit a guilty person. What principles should the court apply if it considers the first of these mistakes to be more serious than the second?

A committee has to make a decision, but its members have different opinions. – What rules should they use to ensure that they can reach a conclusion even if they are in disagreement?

Almost everything that a human being does involves decisions. Therefore, to theorize about decisions is almost the same as to theorize about human

activities. However, decision theory is not quite as all-embracing as that. It focuses on only some aspects of human activity. In particular, it focuses on how we use our freedom. In the situations treated by decision theorists, there are options to choose between, and we choose in a non-random way. Our choices, in these situations, are goal-directed activities. Hence, decision theory is concerned with *goal-directed behaviour in the presence of options*.

We do not decide continuously. In the history of almost any activity, there are periods in which most of the decision-making is made, and other periods in which most of the implementation takes place. Decision-theory tries to throw light, in various ways, on the former type of period.

1.2 A truly interdisciplinary subject

Modern decision theory has developed since the middle of the 20th century through contributions from several academic disciplines. Although it is now clearly an academic subject of its own right, decision theory is typically pursued by researchers who identify themselves as economists, statisticians, psychologists, political and social scientists or philosophers. There is some division of labour between these disciplines. A political scientist is likely to study voting rules and other aspects of collective decision-making. A psychologist is likely to study the behaviour of individuals in decisions, and a philosopher the requirements for rationality in decisions. However, there is a large overlap, and the subject has gained from the variety of methods that researchers with different backgrounds have applied to the same or similar problems.

1.3 Normative and descriptive theories

The distinction between *normative* and *descriptive* decision theories is, in principle, very simple. A normative decision theory is a theory about how decisions should be made, and a descriptive theory is a theory about how decisions are actually made.

The "should" in the foregoing sentence can be interpreted in many ways. There is, however, virtually complete agreement among decision scientists that it refers to the prerequisites of rational decision-making. In other words, a normative decision theory is a theory about how decisions should be made in order to be rational.

This is a very limited sense of the word "normative". Norms of rationality are by no means the only – or even the most important – norms that one may wish to apply in decision-making. However, it is practice to regard norms other than rationality norms as external to decision theory. Decision theory does not, according to the received opinion, enter the scene until the ethical or political norms are already fixed. It takes care of those normative issues that remain even after the goals have been fixed. This remainder of normative issues consists to a large part of questions about how to act in when there is uncertainty and lack of information. It also contains issues about how an individual can coordinate her decisions over time and of how several individuals can coordinate their decisions in social decision procedures.

If the general wants to win the war, the decision theorist tries to tell him how to achieve this goal. The question whether he should at all try to win the war is not typically regarded as a decision-theoretical issue. Similarly, decision theory provides methods for a business executive to maximize profits and for an environmental agency to minimize toxic exposure, but the basic question whether they should try to do these things is not treated in decision theory.

Although the scope of the "normative" is very limited in decision theory, the distinction between normative (i.e. rationality-normative) and descriptive interpretations of decision theories is often blurred. It is not uncommon, when you read decision-theoretical literature, to find examples of disturbing ambiguities and even confusions between normative and descriptive interpretations of one and the same theory.

Probably, many of these ambiguities could have been avoided. It must be conceded, however, that it is more difficult in decision science than in many other disciplines to draw a sharp line between normative and descriptive interpretations. This can be clearly seen from consideration of what constitutes a falsification of a decision theory.

It is fairly obvious what the criterion should be for the falsification of a descriptive decision theory.

- (F1) A decision theory is *falsified as a descriptive theory* if a decision problem can be found in which most human subjects perform in contradiction to the theory.

Since a normative decision theory tells us how a rational agent should act, falsification must refer to the dictates of rationality. It is not evident, however, how strong the conflict must be between the theory and rational decision-making for the theory to be falsified. I propose, therefore, the following two definitions for different strengths of that conflict.

(F2) A decision theory is *weakly falsified as a normative theory* if a decision problem can be found in which an agent can perform in contradiction with the theory without being irrational.

(F3) A decision theory is *strictly falsified as a normative theory* if a decision problem can be found in which an agent who performs in accordance with the theory cannot be a rational agent.

Now suppose that a certain theory T has (as is often the case) been proclaimed by its inventor to be valid both as a normative and as a descriptive theory. Furthermore suppose (as is also often the case) that we know from experiments that in decision problem P , most subjects do not comply with T . In other words, suppose that (F1) is satisfied for T .

The beliefs and behaviours of decision theoreticians are not known to be radically different from those of other human beings. Therefore it is highly probable that at least some of them will have the same convictions as the majority of the experimental subjects. Then they will claim that (F2), and perhaps even (F3), is satisfied. We may, therefore, expect descriptive falsifications of a decision theory to be accompanied by claims that the theory is invalid from a normative point of view. Indeed, this is what has often happened.

1.4 Outline of the following chapters

In chapter 2, the structure of decision processes is discussed. In the next two chapters, the standard representation of decisions is introduced. With this background, various decision-rules for individual decision-making are introduced in chapters 5-10. A brief introduction to the theory of collective decision-making follows in chapter 11.

2. Decision processes

Most decisions are not momentary. They take time, and it is therefore natural to divide them into phases or stages.

2.1 Condorcet

The first general theory of the stages of a decision process that I am aware of was put forward by the great enlightenment philosopher Condorcet (1743-1794) as part of his motivation for the French constitution of 1793. He divided the decision process into three stages. In the first stage, one “discusses the principles that will serve as the basis for decision in a general issue; one examines the various aspects of this issue and the consequences of different ways to make the decision.” At this stage, the opinions are personal, and no attempts are made to form a majority. After this follows a second discussion in which “the question is clarified, opinions approach and combine with each other to a small number of more general opinions.” In this way the decision is reduced to a choice between a manageable set of alternatives. The third stage consists of the actual choice between these alternatives. (Condorcet, [1793] 1847, pp. 342-343)

This is an insightful theory. In particular, Condorcet's distinction between the first and second discussion seems to be a very useful one. However, his theory of the stages of a decision process was virtually forgotten, and does not seem to have been referred to in modern decision theory.

2.2 Modern sequential models

Instead, the starting-point of the modern discussion is generally taken to be John Dewey's ([1910] 1978, pp. 234-241) exposition of the stages of problem-solving. According to Dewey, problem-solving consists of five consecutive stages: (1) a felt difficulty, (2) the definition of the character of that difficulty, (3) suggestion of possible solutions, (4) evaluation of the suggestion, and (5) further observation and experiment leading to acceptance or rejection of the suggestion.

Herbert Simon (1960) modified Dewey's list of five stages to make it suitable for the context of decisions in organizations. According to Simon,

decision-making consists of three principal phases: "finding occasions for making a decision; finding possible courses of action; and choosing among courses of action."(p. 1) The first of these phases he called *intelligence*, "borrowing the military meaning of intelligence"(p. 2), the second *design* and the third *choice*.

Another influential subdivision of the decision process was proposed by Brim et al. (1962, p. 9). They divided the decision process into the following five steps:

1. Identification of the problem
2. Obtaining necessary information
3. Production of possible solutions
4. Evaluation of such solutions
5. Selection of a strategy for performance

(They also included a sixth stage, implementation of the decision.)

The proposals by Dewey, Simon, and Brim et al are all *sequential* in the sense that they divide decision processes into parts that always come in the same order or sequence. Several authors, notably Witte (1972) have criticized the idea that the decision process can, in a general fashion, be divided into consecutive stages. His empirical material indicates that the "stages" are performed in parallel rather than in sequence.

"We believe that human beings cannot gather information without in some way simultaneously developing alternatives. They cannot avoid evaluating these alternatives immediately, and in doing this they are forced to a decision. This is a package of operations and the succession of these packages over time constitutes the total decision-making process." (Witte 1972, p. 180.)

A more realistic model should allow the various parts of the decision process to come in different order in different decisions.

2.3 Non-sequential models

One of the most influential models that satisfy this criterion was proposed by Mintzberg, Raisinghani, and Théorêt (1976). In the view of these authors, the decision process consists of distinct phases, but these phases

do not have a simple sequential relationship. They used the same three major phases as Simon, but gave them new names: identification, development and selection.

The identification phase (Simon's "intelligence") consists of two routines. The first of these is *decision recognition*, in which "problems and opportunities" are identified "in the streams of ambiguous, largely verbal data that decision makers receive" (p. 253). The second routine in this phase is *diagnosis*, or "the tapping of existing information channels and the opening of new ones to clarify and define the issues" (p. 254).

The development phase (Simon's "design") serves to define and clarify the options. This phase, too, consists of two routines. The *search* routine aims at finding ready-made solutions, and the *design* routine at developing new solutions or modifying ready-made ones.

The last phase, the selection phase (Simon's "choice") consists of three routines. The first of these, the *screen* routine, is only evoked "when search is expected to generate more ready-made alternatives than can be intensively evaluated" (p. 257). In the screen routine, obviously suboptimal alternatives are eliminated. The second routine, the *evaluation-choice routine*, is the actual choice between the alternatives. It may include the use of one or more of three "modes", namely (intuitive) judgment, bargaining and analysis. In the third and last routine, *authorization*, approval for the solution selected is acquired higher up in the hierarchy.

The relation between these phases and routines is circular rather than linear. The decision maker "may cycle within identification to recognize the issue during design, he may cycle through a maze of nested design and search activities to develop a solution during evaluation, he may cycle between development and investigation to understand the problem he is solving... he may cycle between selection and development to reconcile goals with alternatives, ends with means". (p. 265) Typically, if no solution is found to be acceptable, he will cycle back to the development phase. (p. 266)

The relationships between these three phases and seven routines are outlined in *diagram 1*.

Exercise: Consider the following two examples of decision processes:

- a. The family needs a new kitchen table, and decides which to buy.

b. The country needs a new national pension system, and decides which to introduce.

Show how various parts of these decisions suit into the phases and routines proposed by Mintzberg et al. Can you in these cases find examples of non-sequential decision behaviour that the models mentioned in sections 2.1-2.2 are unable to deal with?

The decision structures proposed by Condorcet, by Simon, by Mintzberg et al, and by Brim et al are compared in *diagram 2*. Note that the diagram depicts all models as sequential, so that full justice cannot be made to the Mintzberg model.

2.4 The phases of practical decisions – and of decision theory

According to Simon (1960, p. 2), executives spend a large fraction of their time in intelligence activities, an even larger fraction in design activity and a small fraction in choice activity. This was corroborated by the empirical findings of Mintzberg et al. In 21 out of 25 decision processes studied by them and their students, the development phase dominated the other two phases.

In contrast to this, by far the largest part of the literature on decision making has focused on the evaluation-choice routine. Although many empirical decision studies have taken the whole decision process into account, decision *theory* has been exclusively concerned with the evaluation-choice routine. This is "rather curious" according to Mintzberg and coauthors, since "this routine seems to be far less significant in many of the decision processes we studied than diagnosis or design" (p. 257).

This is a serious indictment of decision theory. In its defense, however, may be said that the evaluation-choice routine is the focus of the decision process. It is this routine that makes the process into a *decision* process, and the character of the other routines is to a large part determined by it. All this is a good reason to pay much attention to the evaluation-choice routine. It is not, however, a reason to almost completely neglect the other routines – and this is what normative decision theory is in most cases guilty of.

3. Deciding and valuing

When we make decisions, or choose between options, we try to obtain as good an outcome as possible, according to some standard of what is good or bad.

The choice of a value-standard for decision-making (and for life) is the subject of moral philosophy. Decision theory assumes that such a standard is at hand, and proceeds to express this standard in a precise and useful way.

3.1 Relations and numbers

To see how this can be done, let us consider a simple example: You have to choose between various cans of tomato soup at the supermarket. Your value standard may be related to price, taste, or any combination of these. Suppose that you like soup A better than soup B or soup C, and soup B better than soup C. Then you should clearly take soup A. There is really no need in this simple example for a more formal model.

However, we can use this simple example to introduce two useful formal models, the need for which will be seen later in more complex examples.

One way to express the value pattern is as a *relation* between the three soups: the relation "better than". We have:

A is better than B
B is better than C
A is better than C

Clearly, since A is better than all the other alternatives, A should be chosen.

Another way to express this value pattern is to assign numerical values to each of the three alternatives. In this case, we may for instance assign to A the value 15, to B the value 13 and to C the value 7. This is a *numerical* representation, or representation in terms of numbers, of the value pattern. Since A has a higher value than either B or C, A should be chosen.

The *relational* and *numerical* representations are the two most common ways to express the value pattern according to which decisions are made.

3.2 The comparative value terms

Relational representation of value patterns is very common in everyday language, and is often referred to in discussions that prepare for decisions. In order to compare alternatives, we use phrases such as "better than", "worse than", "equally good", "at least as good", etc. These are all binary relations, i.e., they relate two entities ("arguments") with each other.

For simplicity, we will often use the mathematical notation " $A > B$ " instead of the common-language phrase "A is better than B".

In everyday usage, betterness and worseness are not quite symmetrical. To say that A is better than B is not exactly the same as to say that B is worse than A. Consider the example of a conductor who discusses the abilities of the two flutists of the orchestra he is conducting. If he says "the second flutist is better than the first flutist", he may still be very satisfied with both of them (but perhaps want them to change places). However, if he says "the second flutist is worse than the first flutist", then he probably indicates that he would prefer to have them both replaced.

Exercise: Find more examples of the differences between "A is better than B" and "B is worse than A".

In common language we tend to use "better than" only when at least one of the alternatives is tolerable and "worse than" when this is not the case. (Halldén 1957, p. 13. von Wright 1963, p. 10. Chisholm and Sosa 1966, p. 244.) There may also be other psychological asymmetries between betterness and worseness. (Tyson 1986. Houston et al 1989) However, the differences between betterness and converse worseness do not seem to have enough significance to be worth the much more complicated mathematical structure that would be required in order to make this distinction. Therefore, in decision theory (and related disciplines), the distinction is ignored (or abstracted from, to put it more nicely). Hence,

$A > B$ is taken to represent "B is worse than A" as well as "A is better than B".¹

Another important comparative value term is "equal in value to" or "of equal value". We can use the symbol \equiv to denote it, hence $A \equiv B$ means that A and B have the same value (according to the standard that we have chosen).

Yet another term that is often used in value comparisons is "at least as good as". We can denote it " $A \geq B$ ".

The three comparative notions "better than" ($>$), "equal in value to" (\equiv) and "at least as good as" (\geq) are essential parts of the formal language of *preference logic*. $>$ is said to represent *preference* or *strong preference*, \geq *weak preference*, and \equiv *indifference*.

These three notions are usually considered to be interconnected according to the following two rules:

- (1) A is better than B if and only if A is at least as good as B but B is not at least as good as A. ($A > B$ if and only if $A \geq B$ and not $B \geq A$)
- (2) A is equally good as B if and only if A is at least as good as B and also B at least as good as A. ($A \equiv B$ if and only if $A \geq B$ and $B \geq A$)

The plausibility of these rules can perhaps be best seen from examples. As an example of the first rule, consider the following two phrases:

"My car is better than your car."

"My car is at least as good as your car, but yours is not at least as good as mine."

The second phrase is much more roundabout than the first, but the meaning seems to be the same.

Exercise: Construct an analogous example for the second rule.

The two rules are mathematically useful since they make two of the three notions ($>$ and \equiv) unnecessary. To define them in terms of \geq simplifies

¹ "Worse is the converse of better, and any verbal idiosyncrasies must be disregarded."
(Brogan 1919, p. 97)

mathematical treatments of preference. For our more intuitive purposes, though, it is often convenient to use all three notions.

There is a vast literature on the mathematical properties of \geq , $>$ and \equiv . Here it will be sufficient to define and discuss two properties that are much referred to in decision contexts, namely completeness and transitivity.

3.3 Completeness

Any preference relation must refer to a set of entities, over which it is defined. To take an example, I have a preference pattern for music, "is (in my taste) better music than". It applies to musical pieces, and not to other things. For instance it is meaningful to say that Beethoven's fifth symphony is better music than his first symphony. It is not meaningful to say that my kitchen table is better music than my car. This particular preference relation has musical pieces as its *domain*.

The formal property of completeness (also called connectedness) is defined for a relation *and* its domain.

The relation \geq is complete if and only if for any elements A and B of its domain, either $A \geq B$ or $B \geq A$.

Hence, for the above-mentioned relation to be complete, I must be able to compare any two musical pieces. For instance, I must either consider the Goldberg variations to be at least as good as Beethoven's ninth, or Beethoven's ninth to be at least as good as the Goldberg variations.

In fact, this particular preference relation of mine is not complete, and the example just given illustrates its incompleteness. I simply do not know if I consider the Goldberg variations to be better than the ninth symphony, or the other way around, or if I consider them to be equally good. Perhaps I will later come to have an opinion on this, but for the present I do not. Hence, my preference relation is *incomplete*.

We can often live happily with incomplete preferences, even when our preferences are needed to guide our actions. As an example, in the choice between three brands of soup, A, B, and C, I clearly prefer A to both B and C. As long as A is available I do not need to make up my mind whether I prefer B to C, prefer C to B or consider them to be of equal

value. Similarly, a voter in a multi-party election can do without ranking the parties or candidates that she does not vote for.

Exercise: Can you find more examples of incomplete preferences?

More generally speaking, we were not born with a full set of preferences, sufficient for the vicissitudes of life. To the contrary, most of our preferences have been acquired, and the acquisition of preferences may cost time and effort. It is therefore to be expected that the preferences that guide decisions are in many cases incapable of being represented by a complete preference relation. Nevertheless, in decision theory preference completeness usually accepted as a simplifying assumption. This is also a standard assumption in applications of preference logic to economics and to social decision theory. In economics it may reflect a presumption that everything can be "measured with the measuring rod of money". (Broome 1978, p. 332)

Following tradition in the subject, preference completeness will mostly be assumed in what follows, but the reader should be aware that it is often a highly problematic assumption.

3.4 Transitivity

To introduce the property of transitivity, let us consider the following example of musical preferences:

Bob: "I think Mozart was a much better composer than Haydn."

Cynthia: "What do you think about Beethoven?"

Bob: "Well, in my view, Haydn was better than Beethoven."

Cynthia: "That is contrary to my opinion. I rate Beethoven higher than Mozart."

Bob: "Well, we quite agree. I also think that Beethoven was better than Mozart."

Cynthia: "Do I understand you correctly? Did you not say that Mozart was better than Haydn and Haydn better than Beethoven?"

Bob: "Yes."

Cynthia: "But does it not follow from this that Mozart was better than Beethoven?"

Bob: "No, why should it?"

Bob's position seems strange. What is strange is that his preferences do not satisfy the property of transitivity.

A (strict) preference relation $>$ is transitive if and only if it holds for all elements A, B, and C of its domain that if $A > B$ and $B > C$, then $A > C$.

Although Bob can probably live on happily with his intransitive (= not transitive) preferences, there is a good reason why we consider such preferences to be strange. This reason is that intransitive preferences are often inadequate to guide actions.

To see this, we only have to transfer the example to a case where a decision has to be made. Suppose that Bob has been promised a CD record. He can have either a record with Beethoven's music, one with Mozart's or one with Haydn's. Furthermore suppose that he likes the Mozart record better than the Haydn record, the Haydn record better than the Beethoven record and the Beethoven record better than the Mozart record.

It seems impossible for Bob to make in this case a decision with which he can be satisfied. If he chooses the Mozart record, then he knows that he would have been more satisfied with the Beethoven record. If he chooses Beethoven, then he knows that Haydn would have satisfied him better. However, choosing Haydn would not solve the problem, since he likes Mozart better than Haydn.

It seems as if Bob has to reconsider his preferences to make them useful to guide his decision.

In decision theory, it is commonly supposed that not only strict preference ($>$) but also weak preference (\geq) and indifference (\equiv) are transitive. Hence, the following two properties are assumed to hold:

A weak preference relation \geq is *transitive* if and only if it holds for all elements A, B, and C of its domain that if $A \geq B$ and $B \geq C$, then $A \geq C$.

An indifference relation \equiv is *transitive* if and only if it holds for all elements A, B, and C of its domain that if $A \equiv B$ and $B \equiv C$, then $A \equiv C$.

These properties are generally considered to be more controversial than the transitivity of strict preference. To see why, let us consider the example of 1000 cups of coffee, numbered C_0, C_1, C_2, \dots up to C_{999} .

Cup C_0 contains no sugar, cup C_1 one grain of sugar, cup C_2 two grains etc. Since I cannot taste the difference between C_0 and C_1 , they are equally good in my taste, $C_0 \equiv C_1$. For the same reason, we have $C_1 \equiv C_2, C_2 \equiv C_3$, etc all the way up to $C_{998} \equiv C_{999}$.

If indifference is transitive, then it follows from $C_0 \equiv C_1$ and $C_1 \equiv C_2$ that $C_0 \equiv C_2$. Furthermore, it follows from $C_0 \equiv C_2$ and $C_2 \equiv C_3$ that $C_0 \equiv C_3$. Continuing the procedure we obtain $C_0 \equiv C_{999}$. However, this is absurd since I can clearly taste the difference between C_0 and C_{999} , and like the former much better. Hence, in cases like this (with insufficient discrimination), it does not seem plausible for the indifference relation to be transitive.

Exercise: Show how the same example can be used against indifference of weak preference.

Transitivity, just like completeness, is a common but problematic assumption in decision theory.

3.5 Using preferences in decision-making

In decision-making, preference relations are used to find the *best* alternative. The following simple rule can be used for this purpose:

- (1) An alternative is (*uniquely*) *best* if and only if it is better than all other alternatives. If there is a uniquely best alternative, choose it.

There are cases in which no alternative is uniquely best, since the highest position is "shared" by two or more alternatives. The following is an example of this, referring to tomato soups:

Soup A and soup B are equally good ($A \equiv B$)
Soup A is better than soup C ($A > C$)
Soup B is better than soup C ($B > C$)

In this case, the obvious solution is to pick one of A and B (no matter which). More generally, the following rule can be used:

- (2) An alternative is (among the) best if and only if it is at least as good as all other alternatives. If there are alternatives that are best, pick one of them.

However, there are cases in which not even this modified rule can be used to guide decision-making. The cyclical preferences (Mozart, Haydn, Beethoven) referred to in section 3.4 exemplify this. As has already been indicated, preferences that violate rationality criteria such as transitivity are often not useful to guide decisions.

3.6 Numerical representation

We can also use numbers to represent the values of the alternatives that we decide between. For instance, my evaluation of the collected works of some modern philosophers may be given as follows:

Bertrand Russell 50
Karl Popper 35
WV Quine 35
Jean Paul Sartre 20
Martin Heidegger 1

It follows from this that I like Russell better than any of the other, etc. It is an easy exercise to derive preference and indifference relations from the numbers assigned to the five philosophers. In general, the information provided by a numerical value assignment is sufficient to obtain a relational representation. Furthermore, the weak preference relation thus obtained is always complete, and all three relations (weak and strict preference and indifference) are transitive.

One problem with this approach is that it is in many cases highly unclear what the numbers represent. There is no measure for "goodness as a philosopher", and any assignment of numbers will appear to be arbitrary.

Of course, there are other examples in which the use of numerical representation is more adequate. In economic theory, for example, willingness to pay is often used as a measure of value. (This is another way

of saying that all values are "translated" into monetary value.) If I am prepared to pay, say \$500 for a certain used car and \$250 for another, then these sums can be used to express my (economic) valuation of the two vehicles.

According to some moral theorists, all values can be reduced to one single entity, *utility*. This entity may or may not be identified with units of human happiness. According to utilitarian moral theory, all moral decisions should, at least in principle, consist of attempts to maximize the total amount of utility. Hence, just like economic theory utilitarianism gives rise to a decision theory based on numerical representation of value (although the units used have different interpretations).

Exercise: Consider again Bob's musical preferences, according to the example of the foregoing section. Can they be a given numerical representation?

3.7 Using utilities in decision-making

Numerically represented values (utilities) are easy to use in decision-making. The basic decision-rule is both simple and obvious:

- (1) Choose the alternative with the highest utility.

However, this rule cannot be directly applied if there are more than two alternatives with maximal value, as in the following example of the values assigned by a voter to three political candidates:

Ms. Anderson 15
Mr. Brown 15
Mr. Carpenter 5

For such cases, the rule has to be supplemented:

- (2) Choose the alternative with the highest utility. If more than one alternative has the highest utility, pick one of them (no matter which).

This is a rule of *maximization*. Most of economic theory is based on the idea that individuals maximize their holdings, as measured in money. Utilitarian moral theory postulates that individuals should maximize the utility resulting from their actions. Some critics of utilitarianism maintain that this is to demand too much. Only saints always do the best. For the rest of us, it is more reasonable to just require that we do good enough. According to this argument, in many decision problems there are levels of utility that are lower than maximal utility but still acceptable. As an example, suppose that John hesitates between four ways of spending the afternoon, with utilities as indicated:

- Volunteer for the Red Cross 50
- Volunteer for Amnesty International 50
- Visit aunt Mary 30
- Volunteer for an anti-abortion campaign -50

According to classical utilitarianism, he must choose one of the two maximal alternatives. According to *satisficing* theory, he may choose any alternative that has sufficient utility. If (just to take an example) the limit is 25 units, three of the options are open to him and he may choose whichever of them that he likes.

One problem with satisficing utilitarianism is that it introduces a new variable (the limit for satisfactoriness) that seems difficult to determine in a non-arbitrary fashion. In decision theory, the maximizing approach is almost universally employed.

4. The standard representation of individual decisions

The purpose of this chapter is to introduce decision matrices, the standard representation of a decision problem that is used in mainstream theory of individual decision-making. In order to do this, we need some basic concepts of decision theory, such as alternative, outcome, and state of nature.

4.1 Alternatives

In a decision we choose between different *alternatives* (options). Alternatives are typically courses of action that are open to the decision-maker at the time of the decision (or that she at least believes to be so).²

The set of alternatives can be more or less well-defined. In some decision problems, it is *open* in the sense that new alternatives can be invented or discovered by the decision-maker. A typical example is my decision how to spend this evening.

In other decision problems, the set of alternatives is *closed*, i.e., no new alternatives can be added. A typical example is my decision how to vote in the coming elections. There is a limited number of alternatives (candidates or parties), between which I have to choose.

A decision-maker may restrict her own scope of choice. When deliberating about how to spend this evening, I may begin by deciding that only two alternatives are worth considering, staying at home or going to the cinema. In this way, I have closed my set of alternatives, and what remains is a decision between the two elements of that set.

We can divide decisions with closed alternative sets into two categories: those with voluntary and those with involuntary closure. In cases of voluntary closure, the decision-maker has herself decided to close

² Weirich (1983 and 1985) has argued that options should instead be taken to be decisions that it is possible for the decision-maker to make, in this case: the decision to bring/not to bring the umbrella. One of his arguments is that we are much more certain about what we can decide than about what we can do. It can be rational to decide to perform an action that one is not at all certain of being able to perform. A good example of this is a decision to quit smoking. (A decision merely to try to quit may be less efficient.)

the set (as a first step in the decision). In cases of involuntary closure, closure has been imposed by others or by impersonal circumstances.

Exercise: Give further examples of decisions with alternative sets that are: (a) open (b) voluntarily closed, and (c) involuntarily closed.

In actual life, open alternative sets are very common. In decision theory, however, alternative sets are commonly assumed to be closed. The reason for this is that closure makes decision problems much more accessible to theoretical treatment. If the alternative set is open, a definitive solution to a decision problem is not in general available.

Furthermore, the alternatives are commonly assumed to be *mutually exclusive*, i.e, such that no two of them can both be realized. The reason for this can be seen from the following dialogue:

Bob: "I do not know what to do tomorrow. In fact, I choose between two alternatives. One of them is to go to professor Schleier's lecture on Kant in the morning. The other is to go to the concert at the concert hall in the evening."

Cynthia: "But have you not thought of doing both?"

Bob: "Yes, I may very well do that."

Cynthia: "But then you have *three* alternatives: Only the lecture, only the concert, or both."

Bob: "Yes, that is another way of putting it."

The three alternatives mentioned by Cynthia are mutually exclusive, since no two of them can be realized. Her way of representing the situation is more elaborate and more clear, and is preferred in decision theory.

Hence, in decision theory it is commonly assumed that the set of alternatives is closed and that its elements are mutually exclusive.

4.2 Outcomes and states of nature

The effect of a decision depends not only on our choice of an alternative and how we carry it through. It also depends on factors outside of the decision-maker's control. Some of these extraneous factors are known, they are the *background information* that the decision-maker has. Others are

unknown. They depend on what other persons will do and on features of nature that are unknown to the decision-maker

As an example, consider my decision whether or not to go to an outdoor concert. The outcome (whether I will be satisfied or not) will depend both on natural factors (the weather) and on the behaviour of other human beings (how the band is going to play).

In decision theory, it is common to summarize the various unknown extraneous factors into a number of cases, called *states of nature*.³ A simple example can be used to illustrate how the notion of a state of nature is used. Consider my decision whether or not to bring an umbrella when I go out tomorrow. The effect of that decision depends on whether or not it will rain tomorrow. The two cases "it rains" and "it does not rain" can be taken as the states of nature in a decision-theoretical treatment of this decision.

The possible *outcomes* of a decision are defined as the combined effect of a chosen alternative and the state of nature that obtains. Hence, if I do not take my umbrella and it rains, then the outcome is that I have a light suitcase and get wet. If I take my umbrella and it rains, then the outcome is that I have a heavier suitcase and do not get wet, etc.

4.3 Decision matrices

The standard format for the evaluation-choice routine in (individual) decision theory is that of a *decision matrix*. In a decision matrix, the alternatives open to the decision-maker are tabulated against the possible states of nature. The alternatives are represented by the rows of the matrix, and the states of nature by the columns. Let us use a decision whether to bring an umbrella or not as an example. The decision matrix is as follows:

	It rains	It does not rain
Umbrella	Dry clothes, heavy suitcase	Dry clothes, heavy suitcase
No umbrella	Soaked clothes, light suitcase	Dry clothes, light suitcase

³ The term is inadequate, since it also includes possible decisions by other persons. Perhaps "scenario" would have been a better word, but since "state of nature" is almost universally used, it will be retained here.

For each alternative and each state of nature, the decision matrix assigns an outcome (such as "dry clothes, heavy suitcase" in our example).

Exercise: Draw a decision matrix that illustrates the decision whether or not to buy a ticket in a lottery.

In order to use a matrix to analyze a decision, we need, in addition to the matrix itself, (1) information about how the outcomes are valued, and (2) information pertaining to which of the states of nature will be realized.

The most common way to represent the values of outcomes is to assign utilities to them. Verbal descriptions of outcomes can then be replaced by utility values in the matrix:

	It rains	It does not rain
Umbrella	15	15
No umbrella	0	18

Mainstream decision theory is almost exclusively devoted to problems that can be expressed in matrices of this type, *utility matrices*. As will be seen in the chapters to follow, most modern decision-theoretic methods require numerical information. In many practical decision problems we have much less precise value information (perhaps best expressed by an incomplete preference relation). However, it is much more difficult to construct methods that can deal effectively with non-numerical information.

4.4 Information about states of nature

In decision theory, utility matrices are combined with various types of information about states of nature. As a limiting case, the decision-maker may know which state of nature will obtain. If, in the above example, I know that it will rain, then this makes my decision very simple. Cases like this, when only one state of nature needs to be taken into account, are called "decision-making under certainty". If you know, for each alternative, what will be the outcome if you choose that alternative, then you act under certainty. If not, then you act under non-certainty.

Non-certainty is usually divided into further categories, such as risk, uncertainty, and ignorance. The *locus classicus* for this subdivision is Knight ([1921] 1935), who pointed out that "[t]he term 'risk', as loosely

used in everyday speech and in economic discussion, really covers two things which, functionally at least, in their causal relations to the phenomena of economic organization, are categorically different". In some cases, "risk" means "a quantity susceptible of measurement", in other cases "something distinctly not of this character". He proposed to reserve the term "uncertainty" for cases of the non-quantifiable type, and the term "risk" for the quantifiable cases. (Knight [1921] 1935, pp. 19-20)

In one of the most influential textbooks in decision theory, the terms are defined as follows:

"We shall say that we are in the realm of decision making under:

(a) *Certainty* if each action is known to lead invariably to a specific outcome (the words prospect, stimulus, alternative, etc., are also used).

(b) *Risk* if each action leads to one of a set of possible specific outcomes, each outcome occurring with a known probability. The probabilities are assumed to be known to the decision maker. For example, an action might lead to this risky outcome: a reward of \$10 if a 'fair' coin comes up heads, and a loss of \$5 if it comes up tails. Of course, certainty is a degenerate case of risk where the probabilities are 0 and 1.

(c) *Uncertainty* if either action or both has as its consequence a set of possible specific outcomes, but where the probabilities of these outcomes are completely unknown or are not even meaningful."

(Luce and Raiffa 1957, p. 13)

These three alternatives are not exhaustive. Many – perhaps most – decision problems fall between the categories of risk and uncertainty, as defined by Luce and Raiffa. Take, for instance, my decision this morning not to bring an umbrella. I did not know the probability of rain, so it was not a decision under risk. On the other hand, the probability of rain was not completely unknown to me. I knew, for instance, that the probability was more than 5 per cent and less than 99 per cent. It is common to use the term "uncertainty" to cover, as well, such situations with partial knowledge of the probabilities. This practice will be followed here. The more strict uncertainty referred to by Luce and Raiffa will, as is also common, be called "ignorance". (Cf. Alexander 1975, p. 365) We then have the following scale of knowledge situations in decision problems:

certainty	deterministic knowledge
risk	complete probabilistic knowledge
uncertainty	partial probabilistic knowledge
ignorance	no probabilistic knowledge

It is common to divide decisions into these categories, decisions "under risk", "under uncertainty", etc. These categories will be used in the following chapters.

In summary, the standard representation of a decision consists of (1) a utility matrix, and (2) some information about to which degree the various states of nature in that matrix are supposed to obtain. Hence, in the case of decision-making under risk, the standard representation includes a probability assignment to each of the states of nature (i.e., to each column in the matrix).

5. Expected utility

The dominating approach to decision-making under risk, i.e. known probabilities, is expected utility (EU). This is no doubt "the major paradigm in decision making since the Second World War" (Schoemaker 1982, p. 529), both in descriptive and normative applications.

5.1 What is expected utility?

Expected utility could, more precisely, be called "probability-weighted utility theory". In expected utility theory, to each alternative is assigned a weighted average of its utility values under different states of nature, and the probabilities of these states are used as weights.

Let us again use the umbrella example that has been referred to in earlier sections. The utilities are as follows:

	It rains	It does not rain
Umbrella	15	15
No umbrella	0	18

Suppose that the probability of rain is .1. Then the expected (probability-weighted) utility of bringing the umbrella is $.1 \times 15 + .9 \times 15 = 15$, and that of not bringing the umbrella is $.1 \times 0 + .9 \times 18 = 16.2$. According to the maxim of *maximizing expected utility* (MEU) we should not, in this case, bring the umbrella. If, on the other hand, the probability of rain is .5, then the expected (probability-weighted) utility of bringing the umbrella is $.5 \times 15 + .5 \times 15 = 15$ and that of not bringing the umbrella is $.5 \times 0 + .5 \times 18 = 9$. In this case, if we want to maximize expected utility, then we should bring the umbrella.

This can also be stated in a more general fashion: Let there be n outcomes, to each of which is associated a utility and a probability. The outcomes are numbered, so that the first outcome has utility u_1 and probability p_1 , the second has utility u_2 and probability p_2 , etc. Then the expected utility is defined as follows:

$$p_1 \times u_1 + p_2 \times u_2 + \dots + p_n \times u_n$$

Expected utility theory is as old as mathematical probability theory (although the phrase "expected utility" is of later origin). They were both developed in the 17th century in studies of parlour-games. According to the *Port-Royal Logic* (1662), "to judge what one ought to do to obtain a good or avoid an evil, one must not only consider the good and the evil in itself, but also the probability that it will or will not happen and view geometrically the proportion that all these things have together." (Arnauld and Nicole [1662] 1965, p. 353 [IV:16])

5.2 Objective and subjective utility

In its earliest versions, expected utility theory did not refer to utilities in the modern sense of the word but to monetary outcomes. The recommendation was to play a game if it increased your expected wealth, otherwise not. The probabilities referred to were objective frequencies, such as can be observed on dice and other mechanical devices.

In 1713 Nicolas Bernoulli (1687-1759) posed a difficult problem for probability theory, now known as the St. Petersburg paradox. (It was published in the proceedings of an academy in that city.) We are invited to consider the following game: A fair coin is tossed until the first head occurs. If the first head comes up on the first toss, then you receive 1 gold coin. If the first head comes up on the second toss, you receive 2 gold coins. If it comes up on the third toss, you receive 4 gold coins. In general, if it comes up on the n 'th toss, you will receive 2^n gold coins.

The probability that the first head will occur on the n 'th toss is $1/2^n$. Your expected wealth after having played the game is

$$1/2 \times 1 + 1/4 \times 2 + \dots + 1/2^n \times 2^{n-1} + \dots$$

This sum is equal to infinity. Thus, according to the maxim of maximizing expected wealth a rational agent should be prepared to pay any finite amount of money for the opportunity to play this game. In particular, he should be prepared to put his whole fortune at stake for one single run of the St. Petersburg game.

In 1738 Daniel Bernoulli (1700-1782, a cousin of Nicholas') proposed what is still the conventional solution to the St. Petersburg puzzle. His basic idea was to replace the maxim of maximizing expected wealth by that of maximizing expected (subjective) utility. The utility

attached by a person to wealth does not increase in a linear fashion with the amount of money, but rather increases at a decreasing rate. Your first \$1000 is more worth to you than is \$1000 if you are already a millionaire. (More precisely, Daniel Bernoulli proposed that the utility of the next increment of wealth is inversely proportional to the amount you already have, so that the utility of wealth is a logarithmic function of the amount of wealth.) As can straightforwardly be verified, a person with such a utility function may very well be unwilling to put his savings at stake in the St. Petersburg game.

In applications of decision theory to economic problems, subjective utilities are commonly used. In welfare economics it is assumed that each individual's utility is an increasing function of her wealth, but this function may be different for different persons.

In risk analysis, on the other hand, objective utility is the dominating approach. The common way to measure risk is to multiply "the probability of a risk with its severity, to call that the expectation value, and to use this expectation value to compare risks." (Bondi 1985, p. 9)

"The worst reactor-meltdown accident normally considered, which causes 50 000 deaths and has a probability of 10^{-8} /reactor-year, contributes only about two per cent of the average health effects of reactor accidents." (Cohen 1985, p. 1)

This form of expected utility has the advantage of intersubjective validity. Once expected utilities of the type used in risk analysis have been correctly determined for one person, they have been correctly determined for all persons. In contrast, if utilities are taken to be subjective, then intersubjective validity is lost (and as a consequence of this the role of expert advice is much reduced).

5.3 Appraisal of EU

The argument most commonly invoked in favour of maximizing objectivist expected utility is that this is a fairly safe method to maximize the outcome in the long run. Suppose, for instance, that the expected number of deaths in traffic accidents in a region will be 300 per year if safety belts are compulsory and 400 per year if they are optional. Then, if these calculations are correct, about 100 more persons per year will actually be

killed in the latter case than in the former. We know, when choosing one of these options, whether it will lead to fewer or more deaths than the other option. If we aim at reducing the number of traffic casualties, then this can, due to the law of large numbers, safely be achieved by maximizing the expected utility (i.e., minimizing the expected number of deaths).

The validity of this argument depends on the large number of road accidents, that levels out random effects in the long run. Therefore, the argument is not valid for case-by-case decisions on unique or very rare events. Suppose, for instance, that we have a choice between a probability of .001 of an event that will kill 50 persons and the probability of .1 of an event that will kill one person. Here, random effects will not be levelled out as in the traffic belt case. In other words, we do not know, when choosing one of the options, whether or not it will lead to fewer deaths than the other option. In such a case, taken in isolation, there is no compelling reason to maximize expected utility.

Nevertheless, a decision in this case to prefer the first of the two options (with the lower number of expected deaths) may very well be *based on* a reasonable application of expected utility theory, namely if the decision is included in a sufficiently large group of decisions for which a metadecision has been made to maximize expected utility. As an example, a strong case can be made that a criterion for the regulation of chemical substances should be one of maximizing expected utility (minimizing expected damage). The consistent application of this criterion in all the different specific regulatory decisions should minimize the damages due to chemical exposure.

The larger the group of decisions that are covered by such a rule, the more efficient is the levelling-out effect. In other words, the larger the group of decisions, the larger catastrophic consequences can be levelled out. However, there is both a practical and an absolute limit to this effect. The *practical* limit is that decisions have to be made in manageable pieces. If too many issues are lumped together, then the problems of information processing may lead to losses that outweigh any gains that might have been hoped for. Obviously, decisions can be partitioned into manageable bundles in many different ways, and how this is done may have a strong influence on decision outcomes. As an example, the protection of workers against radiation may be given a higher priority if it is grouped together with other issues of radiation than if it is included among other issues of work environment.

The *absolute* limit to the levelling-out effect is that some extreme effects, such as a nuclear war or a major ecological threat to human life, cannot be levelled out even in the hypothetical limiting case in which all human decision-making aims at maximizing expected utility. Perhaps the best example of this is the Pentagon's use of secret utility assignments to accidental nuclear strike and to failure to respond to a nuclear attack, as a basis for the construction of command and control devices. (Paté-Cornell and Neu 1985)

Even in cases in which the levelling-out argument for expected utility maximization is valid, compliance with this principle is not required by rationality. In particular, it is quite possible for a rational agent to refrain from minimizing total damage in order to avoid imposing high-probability risks on individuals.

To see this, let us suppose that we have to choose, in an acute situation, between two ways to repair a serious gas leakage in the machine-room of a chemical factory. One of the options is to send in the repairman immediately. (There is only one person at hand who is competent to do the job.) He will then run a risk of .9 to die due to an explosion of the gas immediately after he has performed the necessary technical operations. The other option is to immediately let out gas into the environment. In that case, the repairman will run no particular risk, but each of 10 000 persons in the immediate vicinity of the plant runs a risk of .001 to be killed by the toxic effects of the gas. The maxim of maximizing expected utility requires that we send in the repairman to die. This is also a fairly safe way to minimize the number of actual deaths. However, it is not clear that it is the only possible response that is rational. A rational decision-maker may refrain from maximizing expected utility (minimizing expected damage) in order to avoid what would be unfair to a single individual and infringe her rights.

It is essential to observe that expected utility maximization is only meaningful in comparisons between options in one and the same decision. Some of the clearest violations of this basic requirement can be found in risks analysis. Expected utility calculations have often been used for comparisons between risk factors that are not options in one and the same decision. Indeed, most of the risks that are subject to regulation have proponents – typically producers or owners – who can hire a risk analyst to make comparisons such as: "You will have to accept that this risk is smaller than that of being struck by lightning", or: "You must accept this

technology, since the risk is smaller than that of a meteorite falling down on your head." Such comparisons can almost always be made, since most risks are "smaller" than other risks that are more or less accepted. Pesticide residues are negligible if compared to natural carcinogens in food. Serious job accidents are in most cases less probable than highway accidents, etc.

There is no mechanism by which natural food carcinogens will be reduced if we accept pesticide residues. Therefore it is not irrational to refuse the latter while accepting that we have to live with the former. In general, it is not irrational to reject A while continuing to live with B that is much worse than A, if A and B are not options to be chosen between in one and the same decision. To the contrary: To the extent that a self-destructive behaviour is irrational, it would be highly irrational to let oneself be convinced by *all* comparisons of this kind. We have to live with some rather large natural risks, and we have also chosen to live with some fairly large artificial risks. If we were to accept, in addition, all proposed new risks that are small in comparison to some risk that we have already accepted, then we would all be dead.

In summary, the normative status of EU maximization depends on the extent to which a levelling-out effect is to be expected. The strongest argument in favour of objectivist EU can be made in cases when a large number of similar decisions are to be made according to one and the same decision rule.

5.4 Probability estimates

In order to calculate expectation values, one must have access to reasonably accurate estimates of objective probabilities. In some applications of decision theory, these estimates can be based on empirically known frequencies. As one example, death rates at high exposures to asbestos are known from epidemiological studies. In most cases, however, the basis for probability estimates is much less secure. In most risk assessments of chemicals, empirical evidence is only indirect, since it has been obtained from the wrong species, at the wrong dose level and often with the wrong route of exposure. Similarly, the failure rates of many technological components have to be estimated with very little empirical support.

The reliability of probability estimates depends on the absence or presence of systematic differences between objective probabilities and

subjective estimates of these probabilities. Such differences are well-known from experimental psychology, where they are described as lack of *calibration*. Probability estimates are (well-)calibrated if "over the long run, for all propositions assigned a given probability, the proportion that is true equals the probability assigned." (Lichtenstein, et al. 1982, pp. 306-307.) Thus, half of the statements that a well-calibrated subject assigns probability .5 are true, as are 90 per cent of those that she assigns probability .9, etc.

Most calibration studies have been concerned with subjects' answers to general-knowledge (quiz) questions. In a large number of such studies, a high degree of overconfidence has been demonstrated. In a recent study, however, Gigerenzer et al. provided suggestive evidence that the overconfidence effect in general knowledge experiments may depend on biases in the selection of such questions. (Gigerenzer et al 1991)

Experimental studies indicate that there are only a few types of predictions that experts perform in a well-calibrated manner. Thus, professional weather forecasters and horse-race bookmakers make well-calibrated probability estimates in their respective fields of expertise. (Murphy and Winkler 1984. Hoerl and Fallin 1974) In contrast, most other types of prediction that have been studied are subject to substantial overconfidence. Physicians assign too high probability values to the correctness of their own diagnoses. (Christensen-Szalanski and Bushyhead 1981) Geotechnical engineers were overconfident in their estimates of the strength of a clay foundation. (Hynes and Vanmarcke 1976) Probabilistic predictions of public events, such as political and sporting events, have also been shown to be overconfident. In one of the more careful studies of general-event predictions, Fischhoff and MacGregor found that as the confidence of subjects rose from .5 to 1.0, the proportion of correct predictions only increased from .5 to .75. (Fischhoff and MacGregor 1982. Cf: Fischhoff and Beyth 1975. Ronis and Yates 1987.)

As was pointed out by Lichtenstein et al., the effects of overconfidence in probability estimates by experts may be very serious.

"For instance, in the Reactor Safety Study (U.S. Nuclear Regulatory Commission, 1975) 'at each level of the analysis a log-normal distribution of failure rate data was assumed with 5 and 95 percentile limits defined'... The research reviewed here suggests that distributions built from assessments of the .05 and .95 fractiles may

be grossly biased. If such assessments are made at several levels of an analysis, with each assessed distribution being too narrow, the errors will not cancel each other but will compound. And because the costs of nuclear-power-plant failure are large, the expected loss from such errors could be enormous." (Lichtenstein et al. 1982, p. 331)

Perhaps surprisingly, the effects of overconfidence may be less serious when experts' estimates of single probabilities are directly communicated to the public than when they are first processed by decision analysts. The reason for this is that we typically overweight small probabilities. (Tversky and Kahneman 1986) In other words, we make "too little" difference (as compared to the expected utility model) between a situation with, say, a .1 % and a 2 % risk of disaster. This has often been seen as an example of human irrationality. However, it may also be seen as a compensatory mechanism that to some extent makes good for the effects of overconfidence. If an overconfident expert estimates the probability of failure in a technological system at .01 %, then it may be more reasonable to behave as if it is higher than .01 % – as the "unsophisticated" public does – than to behave as if it is exactly .01 % – as experts tend to recommend. It must be emphasized that this compensatory mechanism is far from reliable. In particular, it will distort well-calibrated probabilities, such as probabilities that are calculated from objective frequencies.

In summary, subjective estimates of (objective) probabilities are often unreliable. Therefore, no very compelling argument can be made in favour of maximizing EU if only subjective estimates of the probability values are available.

6. Bayesianism

In chapter 5, probabilities were taken to be frequencies or potential frequencies in the physical world. Alternatively, probabilities can be taken to be purely mental phenomena.

Subjective (personalistic) probability is an old notion. As early as in the *Ars conjectandi* (1713) by Jacques Bernoulli (1654-1705, an uncle of Nicolas and Daniel) probability was defined as a degree of confidence that may be different with different persons. The use of subjective probabilities in expected utility theory, was, however, first developed by Frank Ramsey in the 1930's. Expected utility theory with both subjective utilities and subjective probabilities is commonly called *Bayesian decision theory*, or Bayesianism. (The name derives from Thomas Bayes, 1702-1761, who provided much of the mathematical foundations for modern probabilistic inference.)

6.1 What is Bayesianism?

The following four principles summarize the ideas of Bayesianism. The first three of them refer to the subject as a bearer of a set of probabilistic beliefs, whereas the fourth refers to the subject as a decision-maker.

1. The Bayesian subject has a *coherent set of probabilistic beliefs*. By coherence is meant here formal coherence, or compliance with the mathematical laws of probability. These laws are the same as those for objective probability, that are known from the frequencies of events involving mechanical devices like dice and coins.

As a simple example of *incoherence*, a Bayesian subject cannot have both a subjective probability of .5 that it will rain tomorrow and a subjective probability of .6 that it will either rain or snow tomorrow.

In some non-Bayesian decision theories, notably prospect theory (see section 7.2), measures of degree of belief are used that do not obey the laws of probability. These measures are not probabilities (subjective or otherwise). (Schoemaker, 1982, p. 537, calls them "decision weights".)

2. The Bayesian subject has a *complete set of probabilistic beliefs*. In other words, to each proposition (s)he assigns a subjective probability. A Bayesian subject has a (degree of) belief about everything. Therefore, Bayesian decision-making is always decision-making under certainty or

risk, never under uncertainty or ignorance. (From a strictly Bayesian point of view, the distinction between risk and uncertainty is not even meaningful.)

3. When exposed to new evidence, the Bayesian subject *changes his (her) beliefs in accordance with his (her) conditional probabilities.*

Conditional probabilities are denoted $p(\cdot | \cdot)$, and $p(A|B)$ is the probability that A , given that B is true. ($p(A)$ denotes, as usual, the probability that A , given everything that you know.)

As an example, let A denote that it rains in Stockholm the day after tomorrow, and let B denote that it rains in Stockholm tomorrow. Then Bayesianism requires that once you get to know that B is true, you revise your previous estimate of $p(A)$ so that it coincides with your previous estimate of $p(A|B)$. It also requires that all your conditional probabilities should conform with the definition:

$$p(A|B) = p(A \& B) / p(B)$$

According to some Bayesians (notably Savage and de Finetti) there are no further rationality criteria for your choice of subjective probabilities. As long as you change your mind in the prescribed way when you receive new evidence, your choice of initial subjective probabilities is just a matter of personal taste. Other Bayesians (such as Jeffreys and Jaynes) have argued that there is, given the totality of information that you have access to, a unique admissible probability assignment. (The principle of insufficient reason is used to eliminate the effects of lack of information.) The former standpoint is called subjective (personalistic) Bayesianism. The latter standpoint is called objective (or rationalist) Bayesianism since it postulates a subject-independent probability function. However, in both cases, the probabilities referred to are subjective in the sense of being dependent on information that is available to the subject rather than on propensities or frequencies in the material world.

4. Finally, Bayesianism states that the rational agent *chooses the option with the highest expected utility.*

The descriptive claim of Bayesianism is that actual decision-makers satisfy these criteria. The normative claim of Bayesianism is that rational decision-makers satisfy them. In normative Bayesian decision analysis, "the aim is to reduce a D[ecision] M[aker]'s incoherence, and to make the DM approximate the behaviour of the hypothetical rational agent, so that

after aiding he should satisfy M[aximizing] E[xpected] U[tility]." (Freeling 1984, p. 180)

Subjective Bayesianism does not prescribe any particular relation between subjective probabilities and objective frequencies or between subjective utilities and monetary or other measurable values. The character of a Bayesian subject has been unusually well expressed by Harsanyi:

"[H]e simply cannot help acting *as if* he assigned numerical *utilities*, at least implicitly, to alternative possible outcomes of his behavior, and assigned numerical *probabilities*, at least implicitly, to alternative contingencies that may arise, and *as if* he then tried to maximize his expected utility in terms of these utilities and probabilities chosen by him...

Of course,... we may very well decide to choose these utilities and probabilities in a fully *conscious* and *explicit* manner, so that we can make fullest possible use of our conscious intellectual resources, and of the best information we have about ourselves and about the world. But the point is that the basic claim of Bayesian theory does not lie in the suggestion that we *should* make a conscious effort to maximize our expected utility rather, it lies in the mathematical theorem telling us that *if* we act in accordance with a few very important rationality axioms then we *shall* inevitably maximize our expected utility." (Harsanyi 1977, pp. 381-382)

Bayesianism is more popular among statisticians and philosophers than among more practically oriented decision scientists. An important reason for this is that it is much less operative than most other forms of expected utility. Theories based on objective utilities and/or probabilities more often give rise to predictions that can be tested. It is much more difficult to ascertain whether or not Bayesianism is violated.

"In virtue of these technical interpretations [of utility and probability], a genuine counter-example has to present rational preferences that violate the axioms of preference, or equivalently, are such that there are *no* assignments of probabilities and utilities according to which the preferences maximize expected utility. A genuine counter-example cannot just provide some plausible probability and utility assignments and show that because of

attitudes toward risk it is not irrational to form preferences, or make choices, contrary to the expected utilities obtained from these assignments." (Weirich 1986, p. 422)

As we will see below, fairly plausible counter-examples to Bayesianism can be devised. However for most practical decision problems, that have not been devised to be test cases for Bayesianism, it cannot be determined whether Bayesianism is violated or not.

6.2 Appraisal of Bayesianism

Bayesianism derives the plausibility that it has from quite other sources than objectivist EU theory. Its most important source of plausibility is Savage's representation theorem.

In the proof of this theorem, Savage did not use either subjective probabilities or subjective utilities as primitive notions. Instead he introduced a binary weak preference relation \geq between pairs of alternatives ("is at least as good as"). The rational individual is assumed to order the alternatives according to this relation. Savage proposed a set of axioms for \geq that represents what he considered to be reasonable demands on rational decision-making. According to his theorem, there is, for any preference ordering satisfying these axioms: (1) a probability measure p over the states of the world, and (2) a utility measure u over the set of outcomes, such that the individual always prefers the option that has the highest expected utility (as calculated with these probability and utility measures). (Savage 1954)

The most important of these axioms is the *sure-thing principle*. Let A_1 and A_2 be two alternatives, and let S be a state of nature such that the outcome of A_1 in S is the same as the outcome of A_2 in S . In other words, the outcome in case of S is a "sure thing", not depending on whether one chooses A_1 or A_2 . The sure-thing principle says that if the "sure thing" (i.e. the common outcome in case of S) is changed, but nothing else is changed, then the choice between A_1 and A_2 is not affected.

As an example, suppose that a whimsical host wants to choose a dessert by tossing a coin. You are invited to choose between alternatives A and B . In alternative A , you will have fruit in case of heads and nothing in case of tails. In alternative B you will have pie in case of heads and nothing in case of tails. The decision matrix is as follows:

	Heads	Tails
<i>A</i>	fruit	nothing
<i>B</i>	pie	nothing

When you have made up your mind and announced which of the two alternatives you prefer, the whimsical host suddenly remembers that he has some ice-cream, and changes the options so that the decision matrix is now as follows:

	Heads	Tails
<i>A</i>	fruit	icecream
<i>B</i>	pie	icecream

Since only a "sure thing" (an outcome that is common to the two alternatives) has changed between the two decision problems, the sure thing principle demands that you do not change your choice between *A* and *B* when the decision problem is revised in this fashion. If, for instance, you chose alternative *A* in the first decision problem, then you are bound to do so in the second problem as well.

The starting-point of modern criticism of Bayesianism was provided by Allais (1953). He proposed the following pair of decision problems, now known as the Allais paradox:

"(1) *Préférez-vous la situation A à la situation B?*

SITUATION A: *Certitude de recevoir 100 millions.*

SITUATION B: *10 chances sur 100 de gagner 500 millions.*

89 chances sur 100 de gagner 100 millions.

1 chance sur 100 de ne rien gagner.

"(2) *Préférez-vous la situation C à la situation D?*

SITUATION C: *11 chances sur 100 de gagner 100 millions.*

89 chances sur 100 de ne rien gagner.

SITUATION D: *10 chances sur 100 de gagner 500 millions.*

90 chances sur 100 de ne rien gagner."

(Allais 1953, p. 527)

The two problems can be summarized in the following two decision matrices, where the probabilities of the states of nature have been given within square brackets:

	S₁ [.10]	S₂ [.89]	S₃ [.01]
A	100 000 000	100 000 000	100 000 000
B	500 000 000	100 000 000	0

	S₁ [.10]	S₂ [.89]	S₃ [.01]
C	100 000 000	0	100 000 000
D	500 000 000	0	0

Allais reports that most people prefer *A* to *B* and *D* to *C*. This has also been confirmed in several experiments. This response pattern is remarkable since it is incompatible with Bayesianism. In other words, there is no combination of a subjective probability assignment and a subjective utility assignment such that they yield a higher expected utility for *A* than for *B* and also a higher expected utility for *D* than for *C*. The response also clearly violates the sure-thing principle since the two decision problems only differ in *S*₂, that has the same outcome for both alternatives in each decision problem.

Results contradicting Bayesianism have also been obtained with the following example:

Imagine that the U.S. is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed.

First decision problem: Choose between programs *A* and *B*.

If Program *A* is adopted, 200 people will be saved.

If Program *B* is adopted, there is a 1/3 probability that 600 people will be saved, and a 2/3 probability that no people will be saved.

Second decision problem: Choose between programs *C* and *D*.

If Program *C* is adopted, 400 people will die.

If Program *D* is adopted, there is a 1/3 probability that nobody will die, and a 2/3 probability that 600 people will die. (Tversky and Kahneman 1981, p. 453)

A large majority of the subjects (72 %) preferred program *A* to program *B*, and a large majority (78 %) preferred program *D* to program *C*. However, alternatives *A* and *C* have been constructed to be identical, and so have *B* and *D*. *A* and *B* are framed in terms of the number of lives saved, whereas *C* and *D* are framed in terms of the number of lives lost.

"On several occasions we presented both versions to the same respondents and discussed with them the inconsistent preferences evoked by the two frames. Many respondents expressed a wish to remain risk averse in the 'lives saved' version and risk seeking in the 'lives lost' version, although they also expressed a wish for their answers to be consistent. In the persistence of their appeal, framing effects resemble visual illusions more than computational errors."
(Tversky and Kahneman 1986, p. 260)

What normative conclusions can be drawn from these and other experimental contradictions of Bayesianism? This is one of the most contested issues in decision theory. According to Savage (1954, pp. 101-103), such results do not prove that something is wrong with Bayesianism. Instead, they are proof that the decision-making abilities of most human beings are in need of improvement.

The other extreme is represented by Cohen (1982) who proposes "the norm extraction method" in the evaluation of psychological experiments. This method assumes that "unless their judgment is clouded at the time by wishful thinking, forgetfulness, inattentiveness, low intelligence, immaturity, senility, or some other competence-inhibiting factor, all subjects reason correctly about probability: none are programmed to commit fallacies" (p. 251). He does not believe that there is "any good reason to hypothesise that subjects use an intrinsically fallacious heuristic" (p. 270). If intellectually well-functioning subjects tend to decide in a certain manner, then there must be some rational reason for them to do so.

Essentially the same standpoint was taken by Berkeley and Humphreys (1982), who proposed the following ingenious explanation of why the common reaction to the Asian disease problem may very well be rational.

"Here program *A* appears relatively attractive, as it allows the possibility of finding a way of saving more than 200 people: the future states of the world are not described in the cumulatively exhaustive way that is the case for consequences of program *B*. Program *C* does not permit the possibility of human agency in saving more than 200 lives (in fact, the possibility is left open that one might even *lose* a few more), and given the problem structure... this might well account for preference of *A* over *B*, and *D* over *C*." (Berkeley and Humphreys 1982, p. 222).

Bayesians have found ingenious ways of defending their programme against any form of criticism. However, some of this defense may be counterproductive in the sense of detaching Bayesianism from practical decision science.

7. Variations of expected utility

A large number of models for decision-making under risk have been developed, most of which are variations or generalizations of EU theory. Two of the major variations of EU theory are discussed in this chapter.⁴

7.1 Process utilities and regret theory

In EU, an option is evaluated according to the utility that each outcome has irrespectively of what the other possible outcomes are. However, these are not the only values that may influence decision-makers. A decision-maker may also be influenced by a wish to avoid uncertainty, by a wish to gamble or by other wishes that are related to expectations or to the relations between the actual outcome and other possible outcomes, rather than to the actual outcomes as such. Such values may be represented by numerical values, "process utilities" (Sowden 1984). Although process utilities are not allowed in EU theory, "we can say that there is a presumption in favour of the view that it is not irrational to value certainty as such (because this is in accord with ordinary intuition) and that no argument has been presented – and there seems little prospect of such an argument being presented – that would force us to abandon that presumption." (Sowden 1984, p. 311)

A generalized EU theory (GEU) that takes process utilities into account allows for the influence of attitudes towards risk and certainty. In the words of one of its most persistent proponents, "it resolves Allais's and Ellsberg's paradoxes. By making consequences include risk, it makes expected utilities sensitive to the risks that are the source of trouble in these paradoxes, and so brings M[aximization of] E[xpected] U[tility] into agreement with the preferences advanced in them." (Weirich 1986, p. 436. Cf. Tversky 1975, p. 171.)

It has often been maintained that GEU involves double counting of attitudes to risk. (Harsanyi 1977, p. 385, see also Luce and Raiffa 1957, p. 32.) Weirich (1986, pp. 437-438) has shown that this is not necessarily so. Another argument against GEU was put forward forcefully by Tversky:

⁴ For an overview of the almost bewildering variety of models for decision-making under risk the reader is referred to Fishburn (1989).

"Under the narrow interpretation of Allais and Savage which identifies the consequences with the monetary payoffs, utility theory is violated [in Allais's paradox]. Under the broader interpretation of the consequences, which incorporates non-monetary considerations such as regret utility theory remains intact...

In the absence of any constraints, the consequences can always be interpreted so as to satisfy the axioms. In this case, however, the theory becomes empty from both descriptive and normative standpoints. In order to maximize the power and the content of the theory, one is tempted to adopt a restricted interpretation such as the identification of outcomes with monetary payoffs." (Tversky 1975, p. 171)

This line of criticism may be valid against GEU in its most general form, with no limits to the numbers and types of process utilities. However, such limits can be imposed in a way that is sufficiently strict to make the theory falsifiable without losing its major advantages. Indeed, such a theory has been developed under the name of *regret theory*.

Regret theory (Loomes and Sugden 1982, Bell 1982, Sugden 1986) makes use of a two-attribute utility function that incorporates two measures of satisfaction, namely (1) utility of outcomes, as in classical EU, and (2) quantity of regret. By regret is meant "the painful sensation of recognising that 'what is' compares unfavourably with 'what might have been'." The converse experience of a favourable comparison between the two has been called "rejoicing". (Sugden 1986, p. 67)

In the simplest form of regret theory, regret is measured as "the difference in value between the assets actually received and the highest level of assets produced by *other* alternatives". (Bell 1982, p. 963) The utility function has the form $u(x,y)$, where x represents actually received assets and y the difference just referred to. This function can reasonably be expected to be an increasing function of both x and y . (For further mathematical conditions on the function, see Bell 1982.)

Regret theory provides a simple explanation of Allais's paradox. A person who has chosen option B (cf. section 6.2) has, if state of nature S_3 materializes, strong reasons to regret her choice. A subject who has chosen option D would have much weaker reasons to regret her choice in the case of S_3 . When regret is taken into consideration, it seems quite reasonable to prefer A to B and D to C .

Regret theory can also explain how one and the same person may both gamble (risk prone behaviour) and purchase insurance (risk averse behaviour). Both behaviours can be explained in terms of regret-avoidance. "[I]f you think of betting on a particular horse for the next race and then decide not to, it would be awful to see it win at long odds." (Provided that gambling on the horse is something you might have done, i.e. something that was a real option for you. Cf. Sugden 1986, pp. 72-73.) In the same way, seeing your house burn down after you have decided not to insure it would be an occasion for strongly felt regret.

7.2 Prospect theory

Prospect theory was developed by Kahneman and Tversky ([1979] 1988, 1981) to explain the results of experiments with decision problems that were stated in terms of monetary outcomes and objective probabilities. Nevertheless, its main features are relevant to decision-making in general. Prospect theory differs from most other theories of decision-making by being "unabashedly descriptive" and making "no normative claims". (Tversky and Kahneman 1986, p. 272) Another original feature is that it distinguishes between two stages in the decision process.

The first phase, the *editing phase* serves "to organize and reformulate the options so as to simplify subsequent evaluation and choice." (Kahneman and Tversky [1979] 1988, p. 196) In the editing phase, gains and losses in the different options are identified, and they are defined relative to some neutral reference point. Usually, this reference point corresponds to the current asset position, but it can be "affected by the formulation of the offered prospects, and by the expectations of the decision maker".

In the second phase, the *evaluation phase* the options – as edited in the previous phase – are evaluated. According to prospect theory, evaluation takes place as if the decision-maker used two scales. One of these replaces the monetary outcomes given in the problem, whereas the other replaces the objective probabilities given in the problem.

Monetary outcomes (gains and losses) are replaced by a *value function* v . This function assigns to each outcome x a number $v(x)$, which reflects the subjective value of that outcome. In other words, the value function is a function from monetary gains and losses to a measure of subjective utility. The major difference between this value function and

conventional subjective utility is that it is applied to changes – that is gains and losses – rather than to final states. A typical value function is shown in *diagram 3*. As will be seen, it is concave for gains and convex for losses, and it is steeper for losses than for gains.

Since the value function is different for different reference points (amounts of present wealth), it should in principle be treated as a function of two arguments, $v(w, x)$, where w is the present state of wealth. (For a similar proposal, see Bengt Hansson 1975.) However, this complication of the theory can, for many practical situations, be dispensed with, since "the preference order of prospects is not greatly altered by small or even moderate variations in asset position." (Kahneman and Tversky [1979] 1988, p. 200) As an example, most people are indifferent between a 50 per cent chance of receiving 1000 dollars and certainty of receiving some amount between 300 and 400 dollars, in a wide range of asset positions. (In other words, the "certainty equivalent" of a 50 per cent chance of receiving 1000 dollars is between 300 or 400 dollars.)

Objective probabilities are transformed in prospect theory by a function π that is called the *decision weight*. π is an increasing function from and to the set of real numbers between 0 and 1. It takes the place that probabilities have in expected utility theory, but it does not satisfy the laws of probability. It should not be interpreted as a measure of degree of belief. (Kahneman and Tversky [1979] 1988, p. 202) (As an example of how it violates the laws of probability, let A be an event and let \bar{A} be the absence of that event. Then, if q is a probability measure, $q(A) + q(\bar{A}) = 1$. This does not hold for π . Instead, $\pi(p(A)) + \pi(p(\bar{A}))$ is typically less than 1.)

Diagram 4 shows the decision weight as a function of objective probabilities. Two important features of the decision weight function should be pointed out.

First: Probability differences close to certainty are "overweighted". We consider the difference between a 95 per cent chance of receiving \$ 1 000 000 and certainty to receive \$ 1 000 000 as in some sense bigger than the difference between a 50 per cent chance and a 55 per cent chance to the same amount of money. Similarly, a reduction of the probability of leakage from a waste repository from .01 to 0 is conceived of as more important – and perhaps more worth paying for – than a reduction of the probability from, say, .11 to .10. The overweighting of small probabilities can be used to explain why people both buy insurance and buy lottery tickets.

Secondly: The weighting function is undefined in the areas that are very close to zero and unit probabilities.

"[T]he simplification of prospects in the editing phase can lead the individual to discard events of extremely low probability and to treat events of extremely high probability as if they were certain. Because people are limited in their ability to comprehend and evaluate extreme probabilities, highly unlikely events are either ignored or overweighted and the difference between high probability and certainty is either neglected or exaggerated." (Kahneman and Tversky [1979] 1988, pp. 205-206)

Although the originators of prospect theory have "no normative claims", their theory gives us at least two important lessons for normative theory.

The first of these lessons is the importance of the editing phase or the framing of a decision problem. Rationality demands on the framing of a decision problem should be attended to much more carefully than what has in general been done. Secondly, our tendency to either "ignore" or "overweight" small probabilities has important normative aspects.

It would be a mistake to regard overweighting of small probabilities as a sign of irrationality. It is not *a priori* unreasonable to regard the mere fact that a particular type of event is possible as a relevant factor, irrespectively of the probability that such an event will actually occur. One reason for such a standpoint may be that mere possibilities give rise to process utilities. You may, for instance, prefer not to live in a society in which events of a particular type are possible. Then any option in which the probabilities of such an event is above zero will be associated with a negative (process) utility that will have to be taken into account even if no event of that type actually takes place.

8. Decision-making under uncertainty

8.1 Paradoxes of uncertainty

The discussion about the distinction between uncertainty and probability has centred on two paradoxes. One of them is the paradox of ideal evidence. It was discovered by Peirce ([1878] 1932), but the formulation most commonly referred to is that by Popper:

"Let z be a certain penny, and let a be the statement 'the n th (as yet unobserved) toss of z will yield heads'. Within the subjective theory, it may be assumed that the absolute (or prior) probability of the statement a is equal to $1/2$, that is to say,

$$(1) \quad P(a) = 1/2$$

Now let e be some *statistical evidence*; that is to say, a *statistical report*, based upon the observation of thousands or perhaps millions of tosses of z ; and let this evidence e be *ideally favourable* to the hypothesis that z is strictly symmetrical - that it is a 'good' penny, with equidistribution... We then have no other option concerning $P(a,e)$ [the probability of a , given e] than to assume that

$$(2) \quad P(a,e) = 1/2$$

This means that the probability of tossing heads remains unchanged, in the light of the evidence e , for we now have

$$(3) \quad P(a) = P(a,e).$$

But according to the subjective theory, (3) means that e is, on the whole, (absolutely) irrelevant information with respect to a .

Now this is a little startling; for it means, more explicitly, that our so-called '*degree of rational belief in the hypothesis, a , ought to be completely unaffected by the accumulated evidential knowledge, e* '; that the absence of any statistical evidence concerning z justifies precisely the same '*degree of rational belief*' as the weighty evidence

of millions of observations which, *prima facie*, confirm or strengthen our belief." (Popper [1959] 1980, pp. 407-408)

The paradox lends strong support to Peirce's proposal that "to express the proper state of belief, no *one* number but *two* are requisite, the first depending on the inferred probability, the second on the amount of knowledge on which that probability is based." (Peirce [1878] 1932, p. 421)

The other paradox is *Ellsberg's paradox*. It concerns the following decision problem.

"Imagine an urn known to contain 30 red balls and 60 black and yellow balls, the latter in unknown proportion... One ball is to be drawn at random from the urn; the following actions are considered:

I	\$100	\$0	\$0
II	\$0	\$100	\$0

Action I is 'a bet on red,' II is 'a bet on black.' Which do you prefer?

Now consider the following two actions, under the same circumstances:

	<u>30</u>	<u>60</u>	
	Red	Black	Yellow
III	\$100	\$0	\$100
IV	\$0	\$100	\$100

Action III is a 'bet on red or yellow'; IV is a 'bet on black or yellow.' Which of these do you prefer? Take your time!

A very frequent pattern of response is: action I preferred to II, and IV to III. Less frequent is: II preferred to I, and III preferred to IV." (Ellsberg [1961] 1988, p. 255)

The persons who respond according to any of these patterns violate Bayesianism. They "are simply not acting 'as though' they assigned

numerical or even qualitative probabilities to the events in question". (ibid, p. 257) They also violate the sure-thing principle.⁵

Ellsberg concluded that the degree of uncertainty, or, conversely, the reliability of probability estimates, must be taken into account in decision analysis. This idea has been taken up not only by theoreticians but also by some practitioners of applied decision analysis and decision aiding. Risk analysts such as Wilson and Crouch maintain that "it is the task of the risk assessor to use whatever information is available to obtain a number between zero and one for a risk estimate, with as much precision as is possible, together with an estimate of the imprecision." (Wilson and Crouch 1987, p. 267)

8.2 Measures of incompletely known probabilities

The rules that have been proposed for decision-making under uncertainty (partial probability information) all make use of some quantitative expression of partial probability information. In this section, such "measures of uncertainty" will be introduced. Some decision rules that make use of them will be discussed in section 8.3.

There are two major types of measures of incompletely known probabilities. I propose to call them binary and multi-valued measures.

A *binary measure* divides the probability values into two groups, possible and impossible values. In many cases, the set of possible probability values will form a single interval, such as: "The probability of a major earthquake in this area within the next 20 years is between 5 and 20 per cent."

Binary measures have been used by Ellsberg ([1961] 1988), who referred to a set Y^0 of "reasonable" probability judgments. Similarly, Levi (1986) refers to a "permissible" set of probability judgments. Kaplan has summarized the intuitive appeal of this approach as follows:

⁵ Neither do these persons conform with any of the more common maxims for decisions under ignorance. "They are not 'minimaxing', nor are they applying a 'Hurwicz criterion', maximizing a weighted average of minimum pay-off and maximum for each strategy. If they were following any such rules they would have been indifferent between each pair of gambles, since all have identical minima and maxima. Moreover, they are not 'minimaxing regret', since in terms of 'regrets' the pairs I-II and III-IV are identical." (ibid, p. 257)

"As I see it, giving evidence its due requires that you rule out as too high, or too low, only those values of *con* [degree of confidence] which the evidence gives you *reason* to consider too high or too low. As for the values of *con* not thus ruled out, you should remain undecided as to which to assign." (Kaplan, 1983, p. 570)

Multivalued measures generally take the form of a function that assigns a numerical value to each probability value between 0 and 1. This value represents the degree of reliability or plausibility of each particular probability value. Several interpretations of the measure have been used in the literature:

1. *Second-order probability* The reliability measure may be seen as a measure of the probability that the (true) probability has a certain value. We may think of this as the subjective probability that the objective probability has a certain value. Alternatively, we may think of it as the subjective probability, given our present state of knowledge, that our subjective probability would have had a certain value if we had "access to a certain body of information". (Baron 1987, p. 27)

As was noted by Brian Skyrms, it is "hardly in dispute that people have beliefs about their beliefs. Thus, if we distinguish degrees of belief, we should not shrink from saying that people have degrees of belief about their degrees of belief. It would then be entirely natural for a degree-of-belief theory of probability to treat probabilities of probabilities." (Skyrms 1980, p. 109)

In spite of this, the attitude of philosophers and statisticians towards second-order probabilities has mostly been negative, due to fears of an infinite regress of higher-and-higher orders of probability. David Hume, ([1739] 1888, pp. 182-183) expressed strong misgivings against second-order probabilities. According to a modern formulation of similar doubts, "merely an addition of second-order probabilities to the model is no real solution, for how certain are we about *these* probabilities?" (Bengt Hansson 1975, p. 189)

This is not the place for a discussion of the rather intricate regress arguments against second-order probabilities. (For a review that is favourable to second-order probabilities, see Skyrms 1980. Cf. also Sahlin 1983.) It should be noted, however, that similar arguments can also be devised against the other types of measures of incomplete probability

information. The basic problem is that a precise formalization is sought for the lack of precision in a probability estimate.

2. *Fuzzy set membership* In fuzzy set theory, uncertainty is represented by degrees of membership in a set.

In common ("crisp") set theory, an object is either a member or not a member of a given set. A set can be represented by an indicator function (membership function, element function) μ . Let μ_Y be the indicator function for a set Y . Then for all x , $\mu_Y(x)$ is either 0 or 1. If it is 1, then x is an element of Y . If it is 0, then x is not an element of Y .

In fuzzy set theory, the indicator function can take any value between 0 and 1. If $\mu_Y(x) = .5$, then x is "half member" of Y . In this way, fuzzy sets provide us with representations of vague notions. Vagueness is different from randomness.

"We emphasize the distinction between two forms of uncertainty that arise in risk and reliability analysis: (1) that due to the randomness inherent in the system under investigation and (2) that due to the vagueness inherent in the assessor's perception and judgement of that system. It is proposed that whereas the probabilistic approach to the former variety of uncertainty is an appropriate one, the same may not be true of the latter. Through seeking to quantify the imprecision that characterizes our linguistic description of perception and comprehension, fuzzy set theory provides a formal framework for the representation of vagueness." (Unwin 1986, p. 27)

In fuzzy decision theory, uncertainty about probability is taken to be a form of (fuzzy) vagueness rather than a form of probability. Let α be an event about which the subject has partial probability information (such as the event that it will rain in Oslo tomorrow). Then to each probability value between 0 and 1 is assigned a degree of membership in a fuzzy set A . For each probability value p , the value $\mu_A(p)$ of the membership function represents the degree to which the proposition "it is possible that p is the probability of event α occurring" is true. In other words, $\mu_A(p)$ is the *possibility* of the proposition that p is the probability that a certain event will happen. The vagueness of expert judgment can be represented by possibility in this sense, as shown in *diagram 5*. (On fuzzy representations of uncertainty, see also Dubois and Prade 1988.)

The difference between fuzzy membership and second-order probabilities is not only a technical or terminological difference. Fuzziness is a non-statistical concept, and the laws of fuzzy membership are not the same as the laws of probability.

3. *Epistemic reliability* Gärdenfors and Sahlin ([1982] 1988, cf. also Sahlin 1983) assign to each probability a real-valued measure ρ between 0 and 1 that represents the "epistemic reliability" of the probability value in question. The mathematical properties of ρ are kept open.

The different types of measures of incomplete probabilistic information are summarized in *diagram 6*. As should be obvious, a binary measure can readily be derived from a multivalued measure. Let M_1 be the multivalued measure. Then a binary measure M_2 can be defined as follows, for some real number r : $M_2(p) = 1$ if and only if $M_1(p) \geq r$, otherwise $M_2(p) = 0$. Such a reduction to a binary measure is employed by Gärdenfors and Sahlin ([1982] 1988).

A multivalued measure carries more information than a binary measure. This is an advantage only to the extent that such additional information is meaningful. Another difference between the two approaches is that binary measures are in an important sense more operative. In most cases it is a much simpler task to express one's uncertain probability estimate as an interval than as a real-valued function over probability values.

8.3 Decision criteria for uncertainty

Several decision criteria have been proposed for decision-making under uncertainty. Five of them will be presented here.

1. *Maximin expected utility*. According to *the maximin EU rule*, we should choose the alternative such that its lowest possible EU (i.e., lowest according to any possible probability distribution) is as high as possible (maximize the minimal EU).

For each alternative under consideration, a set of expected values can be calculated that corresponds to the set of possible probability distributions assigned by the binary measure. The lowest utility level that is assigned to an alternative by any of the possible probability distributions is called the "minimal expected utility" of that option. The alternative with the largest minimal expected utility should be chosen. This decision rule

has been called *maximin expected utility* (MMEU) by Gärdenfors (1979). It is an extremely prudent – or pessimistic - decision criterion.

2. *Reliability-weighted expected utility*. If a multivalued decision measure is available, it is possible to calculate the weighted average of probabilities, giving to each probability the weight assigned by its degree of reliability. This weighted average can be used to calculate a definite expected value for each alternative. In other words, the reliability-weighted probability is used in the same way as a probability value is used in decision-making under risk. This decision-rule may be called *reliability-weighted expected utility*.

Reliability-weighted expected utility was applied by Howard (1988) in an analysis of the safety of nuclear reactors. However, as can be concluded from the experimental results on Ellsberg's paradox, it is probable that most people would consider this to be an unduly optimistic decision rule.

Several of the most discussed decision criteria for uncertainty can be seen as attempts at compromises between the pessimism of maximin expected utility and the optimism of reliability-weighted expected utility.

3. *Ellsberg's index*. Daniel Ellsberg proposed the use of an optimism-pessimism index to combine maximin expected utility with what is essentially reliability-weighted expected utility. He assumed that there is both a set Y^0 of possible probability distributions and a single probability distribution y^0 that represents the best probability estimate.

"Assuming, purely for simplicity, that these factors enter into his decision rule in linear combination, we can denote by ρ his degree of confidence, in a given state of information or ambiguity, in the estimated distribution [probability] y^0 , which in turn reflects all of his judgments on the relative likelihood of distributions, including judgments of equal likelihood. Let $\min x$ be the minimum expected pay-off to an act x as the probability distribution ranges over the set Y^0 , let $estx$ be the expected pay-off to the act x corresponding to the estimated distribution y^0 .

The simplest decision rule reflecting the above considerations would be: *Associate with each x the index:*

$$\rho \times estx + (1-\rho) \times \min x$$

Choose that act with the highest index." (Ellsberg [1961] 1988, p. 265)

Here, ρ is an index between 0 and 1 that is chosen so as to settle for the degree of optimism or pessimism that is preferred by the decision-maker.

4. *Gärdenfors's and Sahlin's modified MMEU*. Peter Gärdenfors and Nils-Eric Sahlin have proposed a decision-rule that makes use of a measure ρ of epistemic reliability over the set of probabilities. A certain minimum level ρ_0 of epistemic reliability is chosen. Probability distributions with a reliability lower than ρ_0 are excluded from consideration as "not being serious possibilities". (Gärdenfors and Sahlin [1982] 1988, pp. 322-323) After this, the maximin criterion for expected utilities (MMEU) is applied to the set of probability distributions that are serious possibilities.

There are two extreme limiting cases of this rule. First, if all probability distributions have equal epistemic reliability, then the rule reduces to the classical maximin rule. Secondly, if only one probability distribution has non-zero epistemic reliability, then the rule collapses into strict Bayesianism.

5. *Levi's lexicographical test*. Isaac Levi (1973, 1980, 1986) assumes that we have a permissible set of probability distributions and a permissible set of utility functions. Given these, he proposes a series of three lexicographically ordered tests for decision-making under uncertainty. They may be seen as three successive filters. Only the options that pass through the first test will be submitted to the second test, and only those that have passed through the second test will be submitted to the third.

His first test is E-admissibility. An option is E-admissible if and only if there is some permissible probability distribution and some permissible utility function such that they, in combination, make this option the best among all available options.

His second test is P-admissibility. An option is P-admissible if and only if it is E-admissible and it is also best with respect to the preservation of E-admissible options.

"In cases where two or more cognitive options are E-admissible, I contend that it would be arbitrary in an objectionable sense to choose one over the other except in a way which leaves open the opportunity for subsequent expansions to settle the matter as a result of further inquiry... Thus the rule for ties represents an attitude favoring suspension of judgment over arbitrary choice when, in

cognitive decision making, more than one option is E-admissible."
(Levi 1980, pp. 134-135)

His third test is S-admissibility. For an option to be S-admissible it must both be P-admissible and "security optimal" among the P-admissible alternatives with respect to some permissible utility function. Security optimality corresponds roughly to the MMEU rule. (Levi 1980)

Levi notes that "it is often alleged that maximin is a pessimistic procedure. The agent who uses this criterion is proceeding as if nature is against him." However, since he only applies the maximin rules to options that have already passed the tests of E-admissibility and P-admissibility, this does not apply to his own use of the maximin rule. (Levi 1980, p. 149)

The various decisions rules for uncertainty differ in their practical recommendations, and these differences have given rise to a vivid debate among the protagonists of the various proposals. Ellsberg's proposal has been criticized by Levi (1986, pp. 136-137) and by Gärdenfors and Sahlin ([1982] 1988 pp. 327-330). Levi's theory has been criticized by Gärdenfors and Sahlin ([1982] 1988 pp. 330-333 and 1982b, Sahlin 1985). Levi (1982, 1985 p. 395 n.) has in return criticized the Gärdenfors-Sahlin decision rule. Maher (1989) reports some experiments that seem to imply that Levi's theory is not descriptively valid, to which Levi (1989) has replied.

It would take us too far to attempt here an evaluation of these and other proposals for decision-making under uncertainty. It is sufficient to observe that several well-developed proposals are available and that the choice between them is open to debate. The conclusion for applied studies should be that methodological pluralism is warranted. Different measures of incomplete probabilistic information should be used, including binary measures, second-order probabilities and fuzzy measures. Furthermore, several different decision rules should be tried and compared.

9. Decision-making under ignorance

By decision-making under ignorance is commonly meant decision-making when it is known what the possible states of affairs are, but no information about their probabilities is available. This case, "classical ignorance", is treated in section 9.1.

Situations are not uncommon in which information is lacking not only about the probabilities of the possible states of nature, but also about which these states of affairs are. This more severe form of ignorance about the states of nature are treated in section 9.2.

9.1 Decision rules for "classical ignorance"

The following is a variant of the umbrella example that has been used in previous sections: You have participated in a contest on a TV show, and won the big prize: The Secret Journey. You will be taken by airplane to a one week vacation on a secret place. You do not know where that place is, so for all that you know the probability of rain there may be anything from 0 to 1. Therefore, this is an example of decision-making under ignorance. As before, your decision matrix is:

	It rains	It does not rain
Umbrella	Dry clothes, heavy suitcase	Dry clothes, heavy suitcase
No umbrella	Soaked clothes, light suitcase	Dry clothes, light suitcase

Let us first see what we can do with only a preference relation (i.e., with no information about utilities). As before, your preferences are:

Dry clothes, light suitcase
is better than
Dry clothes, heavy suitcase
is better than
Soaked clothes, light suitcase

Perhaps foremost among the decision criteria proposed for decisions under ignorance is the *maximin* rule: For each alternative, we define its *security level* as the worst possible outcome with that alternative. The maximin rule

urges us to choose the alternative that has the maximal security level. In other words, *maximize* the *minimal* outcome. In our case, the security level of "umbrella" is "dry clothes, heavy suitcase", and the security level of "no umbrella" is "soaked clothes, light suitcase". Thus, the maximin rule urges you to bring your umbrella.

The maximin principle was first proposed by von Neumann as a strategy against an intelligent opponent. Wald (1950) extended its use to games against nature.

The maximin rule does not distinguish between alternatives with the same security level. A variant of it, the *lexicographic maximin*, or *leximin* rule, distinguishes between such alternatives by comparison of their second-worst outcomes. If two alternatives have the same security level, then the one with the highest second-worst outcome is chosen. If both the worst and the second-worst outcomes are on the same level, then the third-worst outcomes are compared, etc. (Sen 1970, ch. 9.)

The maximin and leximin rules are often said to represent extreme prudence or pessimism. The other extreme is represented by the *maximax* rule: choose the alternative whose hope level (best possible outcome) is best. In this case, the *hope level* of "umbrella" is "dry clothes, heavy suitcase", and that of "no umbrella" is "dry clothes, light suitcase". A maximaxer will not bring his umbrella.

It is in general "difficult to justify the maximax principle as rational principle of decision, reflecting, as it does, wishful thinking". (Rapoport 1989, p. 57) Nevertheless, life would probably be duller if not at least some of us were maximaxers on at least some occasions.

There is an obvious need for a decision criterion that does not force us into the extreme pessimism of the maximin or leximin rule or into the extreme optimism of the maximax rule. For such criteria to be practicable, we need utility information. Let us assume that we have such information for the umbrella problem, with the following values:

	It rains	It does not rain
Umbrella	15	15
No umbrella	0	18

A middle way between maximin pessimism and maximax optimism is the *optimism-pessimism index*. (It is often called the *Hurwicz α index*, since it was proposed by Hurwicz in a 1951 paper, see Luce and Raiffa 1957, p.

282. However, as was pointed out by Levi 1980, pp. 145-146, GLS Shackle brought up the same idea already in 1949.)

According to this decision criterion, the decision-maker is required to choose an index α between 0 and 1, that reflects his degree of optimism or pessimism. For each alternative A, let $min(A)$ be its security level, i.e. the lowest utility to which it can give rise, and let $max(A)$ be the hope level, i.e., the highest utility level that it can give rise to. The α -index of A is calculated according to the formula:

$$\alpha \times min(A) + (1-\alpha) \times max(A)$$

Obviously, if $\alpha = 1$, then this procedure reduces to the maximin criterion and if $\alpha = 0$, then it reduces to the maximax criterion.

As can easily be verified, in our umbrella example anyone with an index above 1/6 will bring his umbrella.

Utility information also allows for another decision criterion, namely the *minimax regret* criterion as introduced by Savage (1951, p. 59). (It has many other names, including "minimax risk", "minimax loss" and simply "minimax".)

Suppose, in our example, that you did not bring your umbrella. When you arrive at the airport of your destination, it is raining cats and dogs. Then you may feel regret, "I wish I had brought the umbrella". Your degree of regret correlates with the difference between your present utility level (0) and the utility level of having an umbrella when it is raining (15). Similarly, if you arrive to find that you are in a place where it never rains at that time of the year, you may regret that you brought the umbrella. Your degree of regret may similarly be correlated with the difference between your present utility level (15) and the utility level of having no umbrella when it does not rain (18). A *regret matrix* may be derived from the above utility matrix:

	It rains	It does not rain
Umbrella	0	3
No umbrella	15	0

(To produce a regret matrix, assign to each outcome the difference between the utility of the maximal outcome in its column and the utility of the outcome itself.)

The minimax regret criterion advises you to choose the option with the lowest maximal regret (to *minimize maximal regret*), i.e., in this case to bring the umbrella.

Both the maximin criterion and the minimax regret criterion are rules for the cautious who do not want to take risks. However, the two criteria do not always make the same recommendation. This can be seen from the following example. Three methods are available for the storage of nuclear waste. There are only three relevant states of nature. One of them is *stable rock*, the other is a *geological catastrophe* and the third is *human intrusion* into the depository. (For simplicity, the latter two states of affairs are taken to be mutually exclusive.) To each combination of depository and state of nature, a utility level is assigned, perhaps inversely correlated to the amount of human exposure to ionizing radiation that will follow:

	Stable rock	Geological catastrophe	Human intrusion
Method 1	-1	-100	-100
Method 2	0	-700	-900
Method 3	-20	-50	-110

It will be seen directly that the maximin criterion recommends method 1 and the maximax criterion method 2. The regret matrix is as follows:

	Stable rock	Geological catastrophe	Human intrusion
Method 1	1	50	0
Method 2	0	650	800
Method 3	20	0	10

Thus, the minimax regret criterion will recommend method 3.

A quite different, but far from uncommon, approach to decision-making under ignorance is to try to reduce ignorance to risk. This can (supposedly) be done by use of the *principle of insufficient reason*, that was first formulated by Jacques Bernoulli (1654-1705). This principle states that if there is no reason to believe that one event is more likely to occur than another, then the events should be assigned equal probabilities. The principle is intended for use in situations where we have an exhaustive list of alternatives, all of which are mutually exclusive. In our umbrella example, it leads us to assign the probability $1/2$ to rain.

One of the problems with this solution is that it is extremely dependent on the partitioning of the alternatives. In our umbrella example, we might divide the "rain" state of nature into two or more substates, such as "it rains a little" and "it rains a lot". This simple reformulation reduces the probability of no rain from 1/2 to 1/3. To be useful, the principle of insufficient reason must be combined with symmetry rules for the structure of the states of nature. The basic problem with the principle of insufficient reason, *viz.*, its arbitrariness, has not been solved. (Seidenfeld 1979. Harsanyi 1983.)

The decision rules discussed in this section are summarized in the following table:

Decision rule	Value information needed	Character of the rule
maximin	preferences	pessimism
leximin	preferences	pessimism
maximax	preferences	optimism
optimism-pessimism index	utilities	varies with index
minimax regret	utilities	cautiousness
insufficient reason	utilities	depends on partitioning

The major decision rules for ignorance.

9.2 Unknown possibilities

The case that we discussed in the previous section may also be called *decision-making under unknown non-zero probabilities*. In this case, we know what the possible outcomes are of the various options, but all we know about their probabilities is that they are non-zero. A still higher level of uncertainty is that which results from ignorance of what the possible consequences are, *i.e.*, *decision-making under unknown possibilities*. In probabilistic language, this is the case when there is some consequence for which we do not know whether its probability, given some option, is zero

or non-zero. However, this probabilistic description does not capture the gist of the matter. The characteristic feature of these cases is that we do not have a complete list of the consequences that should be taken into account.

Unknown possibilities are most disturbing when they can lead to catastrophic outcomes. Catastrophic outcomes can be more or less specified, as can be seen from the following series of possible concerns with genetic engineering:

- unforeseen catastrophic consequences
- emergence of new life-forms, with unforeseen catastrophic consequences
- emergence of new viruses, with unforeseen catastrophic consequences
- emergence of new viruses, that will cause many deaths
- emergence of deadly viruses that spread like influenza viruses
- emergence of modified AIDS viruses that spread like influenza viruses

Even if various specified versions of high-level consequence-uncertainty can be shown to be negligible, the underlying more general uncertainty may remain. For instance, even if we can be certain that genetic engineering cannot lead to the emergence of modified AIDS viruses that spread like influenza, they may lead to some other type of catastrophic event that we are not able to foresee. A historical example can be used to illustrate this:

The constructors of the first nuclear bomb were concerned with the possibility that the bomb might trigger an uncontrolled reaction that would propagate throughout the whole atmosphere. Theoretical calculations convinced them that this possibility could be neglected. (Oppenheimer 1980, p. 227) The group might equally well have been concerned with the risk that the bomb could have some other, not thought-of, catastrophic consequence in addition to its (most certainly catastrophic) intended effect. The calculations could not have laid such apprehensions to rest (and arguably no other scientific argument could have done so either).

The implications of unknown possibilities are difficult to come to grips with, and a rational decision-maker has to strike a delicate balance in the relative importance that she attaches to it. An illustrative example is offered by the debate on the polywater hypothesis, according to which

water could exist in an as yet unknown polymeric form. In 1969, *Nature* printed a letter that warned against producing polywater. The substance might "grow at the expense of normal water under any conditions found in the environment", thus replacing all natural water on earth and destroying all life on this planet. (Donahoe 1969) Soon afterwards, it was shown that polywater is a non-existent entity. If the warning had been heeded, then no attempts would have been made to replicate the polywater experiments, and we might still not have known that polywater does not exist.

In a sense, *any* decision may have catastrophic unforeseen consequences. If far-reaching indirect effects are taken into account, then – given the chaotic nature of actual causation – a decision to raise the pensions of government officials may lead to a nuclear holocaust. Any action whatsoever might invoke the wrath of evil spirits (that *might* exist), thus drawing misfortune upon all of us. Appeal to (selected) high-level uncertainties may stop investigations, foster superstition and hence depreciate our general competence as decision-makers.

On the other hand, there are cases in which it would seem unduly risky to entirely dismiss high-level uncertainties. Suppose, for instance, that someone proposes the introduction of a genetically altered species of earthworm that will displace the common earthworm and that will aerate the soil more efficiently. It would not be unreasonable to take into account the risk that this may have unforeseen negative consequences. For the sake of argument we may assume that all concrete worries can be neutralized. The new species can be shown not to induce more soil erosion, not to be more susceptible to diseases, etc. Still, it would not be irrational to say: "Yes, but there may be other negative effects that we have not been able to think of. Therefore, the new species should not be introduced."

Similarly, if someone proposed to eject a chemical substance into the stratosphere for some good purpose or other, it would not be irrational to oppose this proposal solely on the ground that it may have unforeseeable consequences, and this even if all specified worries can be neutralized.

A rational decision-maker should take the issue of unknown possibilities into account in some cases, but not in others. Due to the vague and somewhat elusive nature of this type of uncertainty, we should not expect to find exact criteria for deciding when it is negligible and when it is not. The following list of four factors is meant to be a basic checklist of aspects to be taken into account in deliberations on the seriousness of unknown possibilities.

1. *Asymmetry of uncertainty*: Possibly, a raise in pensions leads in some unknown way to a nuclear war. Possibly, not raising the pensions leads in some unknown way to a nuclear war. We have no reason why one or the other of these two causal chains should be more probable, or otherwise more worth our attention than the other. On the other hand, the introduction of a new species of earthworm is connected with much more consequence-uncertainty than the option not to introduce the new species. Such asymmetry is a necessary but insufficient condition for the issue of unknown possibilities to be non-negligible.

2. *Novelty*: Unknown possibilities come mainly from new and untested phenomena. The emission of a new substance into the stratosphere constitutes a qualitative novelty, whereas an increase in government pensions does not.

An interesting example of the novelty factor can be found in particle physics. Before new and more powerful particle accelerators have been built, physicists have sometimes feared that the new levels of energy may generate a new phase of matter that accretes every atom of the earth. The decision to regard these and similar fears as groundless has been based on observations showing that the earth is already under constant bombardment from outer space of particles with the same or higher energies. (Ruthen 1993)

3. *Spatial and temporal limitations*: If the effects of a proposed measure are known to be limited in space or in time, then these limitations reduce the uncertainty associated with the measure. The absence of such limitations contributes to the relevance of unknown possibilities in many ecological issues, such as global emissions and the spread of chemically stable pesticides.

4. *Interference with complex systems in balance*: Complex systems such as ecosystems and the atmospheric system are known to have reached some type of balance, that may be impossible to restore after a major disturbance. Due to this irreversibility, uncontrolled interference with such systems is connected with serious uncertainty. The same can be said of uncontrolled interference with economic systems. This is an argument for piecemeal rather than drastic economic reforms.

As was mentioned above, the serious cases of unknown possibilities are asymmetrical in the sense that there is at least one option in which this uncertainty is avoided. Therefore, the choice of a strategy-type for the serious cases is much less abstruse than the selection of these cases: The

obvious solution is to avoid the options that are connected to the higher degrees of uncertainty.

This strategy will in many cases take the form of *non-interference* with ecosystems and other well-functioning systems that are insufficiently understood. Such non-interference differs from general conservatism in being limited to a very special category of issues, that does not necessarily coincide with the concerns of political conservatism.

10. The demarcation of decisions

Any analysis of a decision must start with some kind of demarcation of the decision. It must be made clear what the decision is about and what the options are that should be evaluated and chosen between. In practical decision-making, the demarcation is often far from settled. We can distinguish between two degrees of uncertainty of demarcation.

In the first form, the general purpose of the decision is well-determined, but we do not know that all available options have been identified. We can call this decision-making with an *unfinished list of alternatives*. In the second, stronger form, it is not even clear what the decision is all about. It is not well-determined what is the scope of the decision, or what problem it is supposed to solve. This can be called decision-making with an *indeterminate decision horizon*.

10.1 Unfinished list of alternatives

The nuclear waste issue provides a good example of decision-making with an *unfinished list of alternatives*. Perhaps the safest and most economical way to dispose of nuclear waste is yet unknown. Perhaps it will be discovered the year after the waste has been buried in the ground.

There are at least three distinct methods to cope with an unfinished list of options. The first of these is to content oneself with the available list of options, and to choose one of them. In our example, this means that one of the known technologies for nuclear waste disposal is selected to be realized, in spite of the fact that better methods may become available later on. This will be called *closure* of the decision problem.

A second way to cope with an unfinished list of options is to postpone the decision, and search for better options. This will be called *active postponement* (in contrast to "passive postponement", in which no search for more options takes place). In the case of nuclear waste, active postponement amounts to keeping the waste in temporary storage while searching for improved methods of permanent storage.

A third way out is to select and carry out one of the available options, but search for a new and better option and plan for later reconsideration of the issue. For this to be meaningful, the preliminary decision has to be reversible. This will be called *semi-closure* of the

decision. In our example, semi-closure means to select and carry out some method for the possibly final disposal of nuclear waste, such that later retrieval and redisposal of the waste is possible.

The three strategy-types can be summarized as follows:

issue not kept open	issue kept open	
closure	semi-closure	active postponement
something is done now		nothing is done now

The choice between these strategy-types is an integrated part of the overall decision, and it cannot in general be made prior to the actual decision. The division of options into the three strategy-types is an *aspect* of the individual decision. Some of the ways in which this aspect can influence the decision-outcome are summarized in the following five questions:

Do all available alternatives have serious drawbacks? If so, then this speaks against closure.

Does the problem to be solved aggravate with time? If so, then this speaks against active postponement.

Is the best among the reversible alternatives significantly worse than the best among all the alternatives? If so, then this speaks against semi-closure.

Is the search for new alternatives costly? If so, then this speaks against active postponement and semi-closure.

Is there a substantial risk that a decision to search for new alternatives will not be followed through? If so, then this speaks against semi-closure and – in particular – against active postponement.

10.2 Indeterminate decision horizons

Decision-theoretical models presuppose what Savage called a "small world" in which all terminal states (outcomes) are taken to have a definite utility. (Savage 1954) In practice, this is always an idealization. The terminal states of almost all decisions are beginnings of possible future

decision problems. When formulating a decision problem, one has to draw the line somewhere, and determine a "horizon" for the decision. (Toda 1976) There is not in general a single horizon that is the "right" one. In controversial issues, it is common for different interest groups to draw the line differently.

A decision horizon includes a time perspective. We do not plan indefinitely into the future. Some sort of an informal time limit is needed. Unfortunately, the choice of such a limit is often quite arbitrary. In some cases, the choice of a time limit (although mostly not explicit) has a major influence on the decision. A too short perspective can trap the individual into behaviour that she does not want to continue. ("I am only going to smoke this last cigarette".) On the other hand, too long time perspectives make decision-making much too complex.

Nuclear waste provides a good example of the practical importance of the choice of a decision horizon. In the public debate on nuclear waste there are at least four competing decision horizons:

1. The waste disposal horizon: Given the nuclear reactors that we have, how should the radioactive waste be safely disposed of?
2. The energy production horizon: Given the system that we have for the distribution and consumption of energy, how should we produce energy? What can the nuclear waste issue teach us about that?
3. The energy system horizon: Given the rest of our social system, how should we produce, distribute and consume energy? What can the nuclear waste issue teach us about that?
4. The social system horizon: How should our society be organized? What can the nuclear waste issue teach us about that?

Nuclear waste experts tend to prefer the waste disposal horizon. The nuclear industry mostly prefers the energy production horizon, whereas environmentalists in general prefer the energy system horizon or the social system horizon. Each of the four decision horizons for the nuclear waste issue is compatible with rational decision-making. Therefore, different rational decision-makers may have different opinions on what this issue is really about.

Although this is an unusually clear example, it is not untypical of ecological issues. It is common for one and the same environmental problem to be seen by some parties as an isolated problem and by others

merely as part of a more general problem of lifestyle and sustainable development. Whereas some of us see our own country's environmental problems in isolation, others refuse to discuss them on anything but a global scale. This difference in perspective often leads to a divergence of practical conclusions. Some proposed solutions to environmental problems in the industrialized world tend to transfer the problems to countries in the third world. The depletion of natural resources currently caused by the lifestyles of the richer countries would be disastrous if transferred on a per capita basis to a world scale, etc.

Proponents of major social changes are mostly in favour of wide decision horizons. Defenders of the *status quo* typically prefer much narrower decision horizons that leave no scope to radical change. Professional decision analysts also have a predilection for narrow horizons, but at least in part for a different reason: Analytical tools such as mathematical models are in general more readily applicable to decisions with narrow horizons.

There are at least two major strategy-types that can be used to come to grips with this type of uncertainty. One is *subdivision of the decision*. In our example, to achieve this we would have to promote general acceptance that there are several decisions to be made in connection with nuclear waste, each of which requires a different horizon. Each of the four perspectives on nuclear waste is fully legitimate. Everybody, including environmentalists, should accept that we already have considerable amounts of nuclear waste that must be taken care of somehow. To declare that the task is impossible is not much of an option in the context of that decision. On the other hand, everybody, including the nuclear industry, must accept that the nuclear waste issue is also part of various larger social issues. Problems connected to nuclear waste are legitimate arguments in debates on the choice of an energy system and even in debates on the choice of an overall social system.

The other strategy-type is *fusion* of all the proposed horizons, in other words the choice of the narrowest horizon that comprises all the original ones. The rationale for this is that if we have to settle for only one decision horizon, then we should choose one that includes all the considerations that some of us wish to include. In a rational discourse, arguments should not be dismissed merely because they require a widened decision horizon.

From the point of view of rational decision-making, it is much easier to defend a wide decision horizon than a narrow one. Suppose, for instance, that the energy production system is under debate. If the perspective is widened to the energy system as a whole, then we may discover that the best solution to some problems of the production system involves changes in other sectors of the energy system. (E.g., saving or more efficient use of energy may be a better option than any available means of producing more energy.)

On the other hand, our cognitive limitations make wide decision horizons difficult to handle. Therefore, if smaller fractions of the decision-problem can be isolated in a non-misleading way, then this should be done. In social practice, the best criterion for whether or not a subdivision is non-misleading is whether or not it can be agreed upon by all participants in the decision. Therefore, I propose the following rule of thumb for the choice of decision horizons:

- (1) If possible, find a subdivision of the decision-problem that all parties can agree upon. (*subdivision*)
- (2) If that is not possible, settle for the narrowest horizon that includes all the original horizons. (*fusion*)

11. Decision instability

A decision is unstable if the very fact that it has been made provides a sufficient reason to change it. Decision instability has been at the focus of some of the most important developments in decision theory in recent years. After the necessary background has been given in sections 11.1 and 11.2, decision instability will be introduced in section 11.3.

11.1 Conditionalized EU

Let us consider a student who has to decide whether or not to study her textbook before going to an exam. She assigns 10 utility units to passing the exam, and -5 units to reading the textbook. Her situation is covered by the following decision matrix:

	Passes the exam	Does not pass the exam
Studies the textbook	5	-5
Does not study the textbook	10	0

Whether she passes the exam or not, the utility of the outcome will be greater if she has not studied the textbook. It can easily be shown that whatever probability she assigns to passing the exam, the (plain) expected utility of the alternative not to study the textbook is greater than that of studying it. Still, we are not (at least some of us are not) satisfied with this conclusion. The problem is that the probability of passing the exam seems to be influenced by what decision she makes.

In EU theory this problem is solved by conditionalizing of the probabilities that are used in the calculation of expected utility (Jeffrey 1965). In the above example, let " t " stand for the option to read the textbook and " $\neg t$ " for the option not to read it. Furthermore, let " e " stand for the outcome of passing the exam and " $\neg e$ " for that of not passing the exam. Let p be the probability function and u the utility function. Then the unconditional theory gives the following expected utilities:

$$\text{For } t: 5 \times p(e) - 5 \times p(\neg e)$$

$$\text{For } \neg t: 10 \times p(e)$$

This unconditional version of expected utility theory is generally regarded to be erroneous. The correct Bayesian calculation makes use of conditionalized probabilities, as follows: ($p(elt)$ stands for "the probability of e , given that t is true".)

For t : $5 \times p(elt) - 5 \times p(\neg elt)$

For $\neg t$: $10 \times p(e|\neg t)$

It is easy to show that with appropriate conditional probabilities, the expected utility of studying the textbook can be greater than that of not studying it. Using the relationship $p(\neg elt) = 1 - p(elt)$ it follows that the expected utility of t is higher than that of $\neg t$ if and only if $p(elt) - p(e|\neg t) > .5$. In other words, our student will, if she maximizes expected utility, study the textbook if and only if she believes that this will increase her chance of passing the exam by at least .5.

The version of expected utility theory that utilizes conditionalized probabilities is called the *maximization of conditional expected utilities* (MCEU).

11.2 Newcomb's paradox

The following paradox, discovered by the physicist Newcomb, was first published by Robert Nozick (1969): In front of you are two boxes. One of them is transparent, and you can see that it contains \$ 1 000. The other is covered, so that you cannot see its contents. It contains either \$ 1 000 000 or nothing. You have two options to choose between. One is to take both boxes, and the other is to take only the covered box. A good predictor, who has infallible (or almost infallible) knowledge about your psyche, has put the million in the covered box if he predicted that you will only take that box. Otherwise, he has put nothing in it.

Let us apply maximized (conditional) expected utility to the problem. If you decide to take both boxes, then the predictor has almost certainly foreseen this and put nothing in the covered box. Your gain is \$ 1000. If, on the other hand, you decide to take only one box, then the predictor has foreseen this and put the million in the box, so that your gain

is \$ 1 000 000. In other words, maximization of (conditionalized) expected utility urges you to take only the covered box.

There is, however, another plausible approach to the problem that leads to a different conclusion. If the predictor has put nothing in the covered box, then it is better to take both boxes than to take only one, since you will gain \$ 1 000 instead of nothing. If he has put the million in the box, then too it is better to take both boxes, since you will gain \$ 1 001 000 instead of \$ 1 000 000. Thus, taking both boxes is better under all states of nature. (It is a *dominating* option.) It seems to follow that you should take both boxes, contrary to the rule of maximization of (conditional) expected utilities.

A related class of problems is referred to as "medical Newcomb's problems". The best-known of these is the "smoker's dream". According to this story, the smoker dreams that there is no causal connection between smoking and lung cancer. Instead, the observed correlation depends on a gene which causes both lung cancer and smoking in its bearers. The smoker, in this dream, does not know if he has the gene or not. Suppose that he likes smoking, but prefers being a non-smoker to taking the risk of contracting lung cancer. According to expected utility theory, he should refrain from smoking. However, from a causal point of view he should (in this dream of his) continue to smoke. (See Price 1986 for a discussion of medical Newcomb problems.)

The two-box strategy in Newcomb's problem maximizes the "real gain" of having chosen an option, whereas the one-box strategy maximizes the "news value" of having chosen an option. Similarly, the dreaming smoker who stops smoking is maximizing the news value rather than the real value. The very fact that a certain decision has been made in a certain way changes the probabilities that have to be taken into account in that decision.

In *causal decision theory*, expected utility calculations are modified so that they refer to real value rather than news value. This is done by replacing conditional probabilities by some formal means for the evaluation, in terms of probabilities, of the causal implications of the different options. Since there are several competing philosophical views of causality, it is no surprise that there are several formulations of causal decision theory. Perhaps the most influential formulation is that by Gibbard and Harper ([1978] 1988).

According to these authors, the probabilities that a decision-maker should consider are probabilities of counterfactual propositions of the form "if I were to do A, then B would happen". Two such counterfactuals are useful in the analysis of Newcomb's problem, namely:

- (N1) If I were to take only the covered box, then there would be a million in the covered box.
- (N2) If I were to take both boxes, then there would be a million in the covered box.

Using \Rightarrow as a symbol for the counterfactual "if... then ...", these probabilities can be written in the form: $p(A \Rightarrow B)$. Gibbard and Harper propose that all formulas $p(B|A)$ in conditional decision theory should be replaced by $p(A \Rightarrow B)$.

In most cases (such as our above example with the exam), $p(B|A) = p(A \Rightarrow B)$. However, when A is a sign of B without being a cause of B , it may very well be that $p(A \Rightarrow B)$ is not equal to $p(B|A)$. Newcomb's problem exemplifies this. The counterfactual analysis provides a good argument to take two boxes. At the moment of decision, (N1) and (N2) have the same value, since the contents of the covered box cannot be influenced by the choice that one makes. It follows that the expected utility of taking two boxes is larger than that of taking only one.

11.3 Instability

Gibbard and Harper have contributed an example in which their own solution to Newcomb's problem does not work. The example is commonly referred to as "death in Damascus"

"Consider the story of the man who met death in Damascus. Death looked surprised, but then recovered his ghastly composure and said, 'I am coming for you tomorrow'. The terrified man that night bought a camel and rode to Aleppo. The next day, death knocked on the door of the room where he was hiding, and said 'I have come for you'.

'But I thought you would be looking for me in Damascus', said the man.

'Not at all', said death 'that is why I was surprised to see you yesterday. I knew that today I was to find you in Aleppo'.

Now suppose the man knows the following. Death works from an appointment book which states time and place; a person dies if and only if the book correctly states in what city he will be at the stated time. The book is made up weeks in advance on the basis of highly reliable predictions. An appointment on the next day has been inscribed for him. Suppose, on this basis, the man would take his being in Damascus the next day as strong evidence that his appointment with death is in Damascus, and would take his being in Aleppo the next day as strong evidence that his appointment is in Aleppo...

If... he decides to go to Aleppo, he then has strong grounds for expecting that Aleppo is where death already expects him to be, and hence it is rational for him to prefer staying in Damascus. Similarly, deciding to stay in Damascus would give him strong grounds for thinking that he ought to go to Aleppo..."(Gibbard and Harper [1978] 1988, pp. 373-374)

Once you know that you have chosen Damascus, you also know that it would have been better for you to choose Aleppo, and *vice versa*. We have, therefore, a case of *decision instability*: whatever choice one makes, the other choice would have been better.

Richter (1984) has proposed a slight modification of the death in Damascus case:

"Suppose the man's mother lives in Damascus but the man takes this fact to provide no independent evidence to Death's being or not being in Damascus that night. Suppose also that the man quite reasonably prefers the outcome of dying in Damascus to that of dying in Aleppo for the simple reason that dying in Damascus would afford him a few last hours to visit his mother. Of course he still prefers going to Aleppo and living to visiting his mother and dying. Now we ought to say in this case that since he can't escape the certainty of death no matter what he does, that rationality ought to require going to Damascus." (Richter 1984, p. 396)

Causal decision theory (the theory that leads us to take both boxes in Newcomb's example) cannot adequately account for rational choice in this example. Although going to Damascus clearly is the most reasonable thing to do, it is not a stable alternative. There is, in this case, simply no alternative that satisfies both of the conditions to be stable and to maximize real value.

In the rapidly expanding literature on decision instability, various attempts at formal explications of instability have been proposed and put to test. Different ways to combine expected utility maximization with stability tests have been proposed. Furthermore, there is an on-going debate on the normative status of stability, i.e., on the issue of whether or not a rational solution to a decision problem must be a stable solution. Some of the most important contributions in the field, besides those already referred to, are papers by Eells (1985), Horwich (1985, p. 445), Rabinowicz (1989), Richter (1986), Skyrms (1982, 1986), Sobel (1990), and Weirich (1985).

12. Social decision theory

Decision rules that have been developed for individual decision-making may in many cases also be used for decision-making by groups. As one example, theories of legal decision-making do not in general make a difference between decisions by a single judge and decisions by several judges acting together as a court of law. The presumption is that the group acts as if it were a single individual. Similarly, most theories for corporate decision-making treat the corporation as if all decisions were to be taken by a single individual decision-maker. (Cf. Freeling 1984, p. 200) Indeed, "[a]ny decision maker - a single human being or an organization - which can be thought of as having a unitary interest motivating its decisions can be treated as an individual in the theory". (Luce and Raiffa 1957, p. 13)

By a collective decision theory is meant a theory that models situations in which decisions are taken by two or more persons, who may have conflicting goals or conflicting views on how the goals should be achieved. Such a theory treats individuals as "having conflicting interests which must be resolved, either in open conflict or by compromise". (Luce and Raiffa 1957, p. 13) Most studies in collective decision theory concern voting, bargaining and other methods for combining individual preferences or choices into collective decisions.

The most important concern of social decision theory is the aggregation of individual preferences (choices) into collective preferences (choices). The central problem is to find, given a set of individual preferences, a rational way to combine them into a set of social preferences or into a social choice.

Social decision theory is not a smaller field of knowledge than individual decision theory. Therefore, this short chapter can only be a very rudimentary introduction.

12.1 The basic insight

The fundamental insight in social decision theory was gained by Borda and Condorcet, but forgotten for many years. They discovered that in simple majority rule, there may be situations in which every option is unstable in the sense that a majority coalition can be formed against it. To see what this means in practice, let us consider the following example.

We will assume that three alternatives are available for the handling of nuclear waste. The nuclear industry has worked out a proposal, and provided documentation to show that it is safe enough. We will call this the "industry proposal". A group of independent scientists, who were sceptical of the industry proposal, developed a proposal of their own. It contains several more barriers than the industry proposal, and is therefore considered to be safer. On the other hand, it is several times more expensive. We will call this the "expensive solution". But in spite of the extra barriers, many environmentalists have not been convinced even by the expensive solution. They propose that the whole issue should be postponed until further studies have been conducted.

In parliament, there are three factions of approximately the same size. The members of the first faction (the "economists") are mostly concerned with economic and technological development. They put the industry proposal first. In the choice between postponement and the expensive solution, they prefer the former, for economic reasons. Thus, their preferences are:

Economists:

1. industry proposal
2. postponement
3. expensive solution

The second faction (the "ethicists") is most of all concerned with our responsibility not to hand over the problem to the generations after us. They want the problem to be solved *now*, with the best method that is available. Their preferences are:

Ethicists:

1. expensive solution
2. industry proposal
3. postponement

The third group (the "environmentalists") prefer to postpone the final deposition of the waste, since they do not believe even in the expensive solution. Their preferences are:

Environmentalists:

1. postponement
2. expensive solution
3. industry proposal

Now let us see what happens in majority voting. First suppose that the industry proposal wins. Then a coalition of ethicists and environmentalists can be formed to change the decision, since these two groups both prefer the expensive solution to the industry proposal.

Next, suppose that the expensive solution has won. Then a coalition to change the decision can be formed by economists and environmentalists, since they both prefer postponement to the expensive solution.

Finally, suppose that postponement has won. Then the decision can be changed by a coalition of economists and ethicists, who both prefer the industry proposal to postponement.

We started with three reasonably rational patterns of individual preferences. We used what we believed to be a rational method for aggregation, and arrived at cyclic social preferences.

12.2 Arrow's theorem

The starting-point of modern social decision theory was a theorem by Kenneth Arrow (1951). He set out to investigate whether there is some other social decision rule than majority rule, under which cyclic social preferences can be avoided. The answer, contained in his famous theorem, is that if four seemingly reasonable rationality criteria are satisfied by the decision rule, then cyclicity cannot be avoided. For an accessible proof of the theorem, the reader is referred to Sen (1970, ch. 3*.).

In the decades that have followed, many more results of a similar nature have accumulated. When the range of alternatives is extended from a simple list, as in our example, to the set of all points in a Euclidean space, still stronger impossibility results than Arrow's can be obtained. (McKelvey 1976, 1979, Schofield 1978) It is characteristic of social decision theory that almost all of its more important results are of a negative nature, showing that some rationality demands on a social decision procedure are not compatible.

References

Alexander, Ernest R (1970), "The limits of uncertainty: A note", *Theory and Decision* 6:363-370.

Allais, M (1953), "Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'école américaine", *Econometrica* 21:503-546.

Alpert, Marc and Howard Raiffa (1982), "A progress report on the training of probability assessors", pp. 294-305 in Kahneman et al. (1982).

Arnauld, Antoine and Pierre Nicole ([1662] 1965), *La logique ou l'art de penser*.

Arrow, Kenneth (1951), *Social choice and individual values*.

Baron, Jonathan (1987), "Second-order probabilities and belief functions", *Theory and Decision* 23:25-36.

Bell, David E (1982), "Regret in Decision Making under Uncertainty", *Operations Research* 30:961-981.

Berkeley, Dina and Patrick Humphreys (1982), "Structuring decision problems and the bias heuristic", *Acta Psychologica* 50:201-252.

Bondi, Herman (1985), "Risk in perspective", pp. 8-17 in MG Cooper (ed.), *Risk*.

Brim, Orville G et al. (1962) *Personality and Decision Processes, Studies in the Social Psychology of Thinking*, Stanford 1962,

Brogan, Albert P (1919) "The Fundamental Value Universal", *Journal of Philosophy, Psychology, and Scientific Methods* 16:96-104.

Broome, John (1978) "Choice and value in economics", *Oxford Economic Papers* 30:313-333.

- Chisholm, Roderick M and Ernest Sosa (1966) "On the Logic of 'Intrinsically Better'", *American Philosophical Quarterly* 3:244-249.
- Christensen-Szalanski, JJJ and JB Bushyhead (1981) "Physicians' use of probabilistic information in a real clinical setting", *Journal of Experimental Psychology: Human Perception and Performance* 7:928-935.
- Cohen, Bernhard L (1983), "Probabilistic Risk Assessment of Wastes Buried in the Ground", *Risk Analysis* 3:237-243.
- Cohen, Bernhard L (1985), "Criteria for Technology Acceptability", *Risk Analysis* 5:1-3.
- Cohen, L Jonathan (1982), "Are People Programmed to Commit Fallacies? Further Thoughts about the Interpretation of Experimental Data on Probability Judgment", *Journal for the Theory of Social Behaviour* 12:251-274.
- Condorcet ([1793] 1847), "Plan de Constitution, présenté a la convention nationale les 15 et 16 février 1793", *Oeuvres*, vol. 12, pp. 333-415.
- Dewey, John ([1910] 1978), *How We Think*, pp. 177-356 in *Middle Works*, vol 6.
- Donahoe, FJ (1969) "Anomalous" Water, *Nature* 224:198.
- Dubois, Didier and Henri Prade (1988), "Decision evaluation methods under uncertainty and imprecision", pp. 48-65 in J Kacprzyk and M Fedrizzi (eds.), *Combining Fuzzy Impression with Probabilistic Uncertainty in Decision Making*.
- Eells, Ellery (1985), "Weirich on decision instability", *Australasian Journal of Philosophy* 63:475-478.
- Ellsberg, Daniel ([1961] 1988), "Risk, ambiguity, and the Savage axioms", pp. 245-269 in Gärdenfors and Sahlin (1988).

Fischhoff, Baruch and R Beyth (1975) "I knew it would happen" remembered probabilities of once-future things", *Organizational Behavior and Human Performance* 13:1-16.

Fischhoff, Baruch and Don MacGregor (1982) "Subjective Confidence in Forecasts", *Journal of Forecasting* 1:155-172.

Fischhoff, Peter C, P Slovic and S Lichtenstein (1977), "Knowing with certainty: The appropriateness of extreme confidence", *Journal of Experimental Psychology: Human Perception and Performance* 3:552-564.

Fishburn, Peter C (1989), "Generalizations of expected utility theories: A survey of recent proposals", *Annals of Operations Research* 19:3-28.

Freeling, Anthony NS (1984), "A philosophical basis for decision aiding", *Theory and Decision* 16:179-206.

Gärdenfors, Peter (1979), "Forecasts, decisions and uncertain probabilities", *Erkenntnis* 14:159-181.

Gärdenfors, Peter and Nils-Eric Sahlin ([1982] 1988), "Unreliable probabilities, risk taking, and decision making", pp. 313-334 in Gärdenfors and Sahlin (1988).

Gärdenfors, Peter and Nils-Eric Sahlin (1982b), "Reply to Levi", *Synthese* 53:433-438.

Gärdenfors, Peter and Nils-Eric Sahlin, eds. (1988), *Decision, probability, and utility*.

Gibbard, Alan and William L Harper ([1978] 1988), "Counterfactuals and two kinds of expected utility", pp. 341-376 in Gärdenfors and Sahlin (1988).

Gigerenzer, Gerd Ulrich Hoffrage, and Heinz Kleinbölting (1991) "Probabilistic mental models", *Psychological Review* 98:506-528.

Halldén, Sören (1957) *On the Logic of 'Better'*.

- Hansson, Bengt (1975), "The appropriateness of the expected utility model", *Erkenntnis* 9:175-193.
- Hansson, Sven Ove (1987), *Risk Decisions and Nuclear Waste*, SKN Report 19, Stockholm.
- Hansson, Sven Ove (1989a), "Dimensions of Risk", *Risk Analysis* 9:107-112.
- Hansson, Sven Ove (1989b), "Preference logic as a model for risk comparisons", pp. 355-364 in *Proceedings, Management of Risk from Genotoxic Substances in the Environment*, Stockholm.
- Hansson, Sven Ove (1989c) "A New Semantical Approach to the Logic of Preference", *Erkenntnis* 31:1-42.
- Hansson, Sven Ove (1992) "A Procedural Model of Voting", *Theory and Decision* 32:269-301.
- Hansson, Sven Ove (1993) "Money-Pumps, Self-Torturers and the Demons of Real Life", *Australasian Journal of Philosophy*, in press.
- Hansson, Sven Ove (1994a) "Changes in Preference", *Theory and Decision*, in press.
- Hansson, Sven Ove (1994b) "What is Ceteris Paribus Preference?", *Journal of Philosophical Logic*, in press.
- Harsanyi, John C (1977), "On the rationale of the Bayesian approach: Comments on professor Watkin's paper", pp. 381-392 in Butts and Hintikka (eds.), *Foundational Problems in the Special Sciences*.
- Harsanyi, John C (1983), "Bayesian decision theory, subjective and objective probabilities, and acceptance of empirical hypotheses", *Synthese* 57:341-365.

Hoerl, Arthur E and Herbert K Fallin (1974), "Reliability of Subjective Evaluations in a High Incentive Situation", *Journal of the Royal Statistical Society* 137, Part 2, 227-230.

Horwich, Paul (1985), "Decision theory in light of Newcomb's problem", *Philosophy of Science* 52:431-450.

Houston, David A et al (1989) "The Influence of Unique Features and Direction of Comparison on Preferences", *Journal of Experimental Social Psychology* 25:121-141.

Howard, Ronald A (1988), "Uncertainty about Probability: A Decision Analysis Perspective", *Risk Analysis* 8:91-98.

Hume, David ([1739] 1888) *A Treatise of Human Nature*, ed. by LA Selby-Bigge.

Hynes, M and E Vanmarcke (1976), "Reliability of embankment performance predictions", *Proceedings of the ASCE Engineering Mechanics Division, Specialty Conference*, Waterloo, Ontario, Canada, University of Waterloo Press.

Jeffrey, Richard C (1965) *The Logic of Decision*.

Kahneman, Daniel and Amos Tversky ([1979] 1988), "Prospect Theory: An analysis of decision under risk", pp. 183-214 in Gärdenfors and Sahlin (1988).

Kahneman, Daniel et al. (1982), *Judgment under uncertainty, heuristics and biases*.

Kaplan, Mark (1983), "Decision theory as philosophy", *Philosophy of Science* 50:549-577.

Knight, FH ([1921] 1935), *Risk, Uncertainty and Profit*.

Levi, Isaac (1973), *Gambling with truth*.

- Levi, Isaac (1980), *The enterprise of knowledge*.
- Levi, Isaac (1982), "Ignorance, probability and rational choice", *Synthese* 53:387-417.
- Levi, Isaac (1985), "Imprecision and indeterminacy in probability judgment", *Philosophy of Science* 52:390-409.
- Levi, Isaac (1986), *Hard Choices: Decision Making under Unresolved Conflict*.
- Levi, Isaac (1989), "Reply to Maher", *Economics and Philosophy* 5:79-90.
- Lichtenstein, Sarah et al. (1982), "Calibration of probabilities: The state of the art to 1980", pp. 306-334 in Kahneman et al. (1982).
- Loomes, G and R Sugden (1982), "Regret theory: an alternative theory of rational choice under uncertainty", *Economic Journal* 92:805-824.
- Luce, R Duncan and Howard Raiffa (1957), *Games and Decisions*.
- Maher, Patrick (1989), "Levi on the Allais and Ellsberg paradoxes", *Economics and Philosophy* 5:69-78.
- McKelvey, Richard D (1976), "Intransitivities in Multidimensional Voting Models and some Implications for Agenda Control", *Journal of Economic Theory* 12:472-482.
- McKelvey, Richard D (1979), "General Conditions for Global Intransitivities in Formal Voting Models", *Econometrica* 47:1085-1112.
- Manski, Charles F (1988), "Ordinal utility models of decision making under uncertainty", *Theory and Decision* 25:79-104.
- Mintzberg, Henry, Dury Raisinghani, and André Théorêt (1976), "The Structure of 'Unstructured' Decision Processes", *Administrative Sciences Quarterly* 21:246-275.

Murphy, AH and RL Winkler (1984) "Probability forecasting in meteorology", *Journal of the American Statistical Association* 79:489-500.

Nozick, Robert (1969), "Newcomb's problem and two principles of choice", pp. 114-146 in N Rescher et al. (eds.), *Essays in Honor of Carl G. Hempel*.

Oppenheimer, Robert (1980) *Letters and Recollections* (ed. Alice Kimball Smith and Charles Weiner), Harvard University Press.

Paté-Cornell, ME and JE Neu (1985) "Warning Systems and Defense Policy: A Reliability Model for the Command and Control of U.S. Nuclear Forces", *Risk Analysis* 5:121-138.

Peirce, CS ([1878] 1932) "The probability of induction", pp. 415-432 in *Collected papers*, volume 2.

Peterson, Cameron R (1965), "Internal consistency of subjective probabilities", *Journal of Experimental Psychology* 70:526-533.

Popper, Karl ([1959] 1980), *The Logic of Scientific Discovery*.

Rabinowicz, Wlodzimierz (1989), "Stable and retrievable options", *Philosophy of Science* 56:624-641.

Rapoport, Anatol (1989), *Decision Theory and Decision Behaviour*.

Richter, Reed (1984), "Rationality revisited", *Australasian Journal of Philosophy* 62:392-403.

Richter, Reed (1986), "Further comments on decision instability", *Australasian Journal of Philosophy* 64:345-349.

Ronis, David L and J Frank Yates (1987) "Components of Probability Judgment Accuracy: Individual Consistency and Effects of Subject Matter and Assessment Method", *Organizational Behavior and Human Decision Processes* 40:193-218.

- Ruthen, Russell (1993) "Strange Matter", *Scientific American* August 1993, p. 10.
- Sahlin, Nils-Eric (1983), "On second order probability and the notion of epistemic risk", pp. 95-104 in BP Stigum and F Wenztop (eds.), *Foundations of Utility Theory with Applications*.
- Sahlin, Nils-Eric (1984). "Level of aspiration and risk", mimeographed, Department of Philosophy, Lund University. Lund.
- Sahlin, Nils-Eric (1985), "Three decision rules for generalized probability representation", *The Behavioral and Brain Sciences* 8:751-753.
- Sahlin, Nils-Eric (1988), "The significance of empirical evidence for developments in the foundations of decision theory", pp. 103-121 in D Batens and JP van Bendegem (eds) *Theory and experiment*.
- Samuelson, William and Richard Zeckhauser (1988), "Status Quo Bias in Decision Making", *Journal of Risk and Uncertainty* 1:7-59.
- Savage, LJ (1951), "The theory of statistical decision", *Journal of the American Statistical Association* 46:55-67.
- Savage, LJ (1954), *The Foundations of Statistics*.
- Schoemaker, Paul JH (1982), "The Expected Utility Model: Its Variants, Purposes, Evidence and Limitations", *Journal of Economic Literature* 20:529-563.
- Schofield, Norman (1978), "Instability of Simple Dynamic Games", *Review of Economic Studies* 45:575-594.
- Seidenfeld, Teddy (1979), "Why I am not an objective Bayesian; Some reflections prompted by Rosenkrantz", *Theory and Decision* 11:413-440.
- Sen, Amartya K (1970), *Collective Choice and Social Welfare*.
- Simon, Herbert (1960), *The New Science of Management Decision*.

Skyrms, Brian (1980), "Higher order degrees of belief", pp. 109-137 in DH Mellor (ed.), *Prospects for Pragmatism*.

Skyrms, Brian (1982), "Causal decision theory", *Journal of Philosophy* 79:695-711.

Skyrms, Brian (1986), "Deliberational Equilibria", *Topoi* 5:59-67.

Sobel, Jordan Howard (1990), "Maximization, stability of decision, and actions in accordance with reason", *Philosophy of Science* 57:60-77.

Sowden, Lanning (1984), "The inadequacy of Bayesian decision theory", *Philosophical Studies* 45:293-313.

Sugden, Robert (1986), "Regret, recrimination and rationality", pp. 67-80 in L Daboni et al. (eds.), *Recent Developments in the Foundations of Utility and Risk Theory*.

Suppes, Patrick (1961), "The philosophical relevance of decision theory", *Journal of Philosophy* 58:605-614.

Toda, Masanao (1976) "The Decision Process: A Perspective", *International Journal of General Systems* 3:79-88.

Tversky, Amos (1975), "A critique of expected utility theory: Descriptive and normative considerations", *Erkenntnis* 9:163-173.

Tversky, Amos and Daniel Kahneman (1981), "The Framings of Decisions and the Psychology of Choice", *Science* 211:453-458.

Tversky, Amos and Daniel Kahneman (1986), "Rational Choice and the Framing of Decisions", *Journal of Business* 59:251-278.

Tyson, PD (1986) "Do your standards make any difference? Asymmetry in preference judgments", *Perceptual and Motor Skills* 63:1059-1066.

Unwin, Stephen D (1986), "A Fuzzy Set Theoretic Foundation for Vagueness in Uncertainty Analysis", *Risk Analysis* 6:27-34.

Wald, A (1950), *Statistical Decision Functions*.

Weirich, Paul (1983), "A decision maker's options", *Philosophical Studies* 44:175-186.

Weirich, Paul (1985), "Decision instability", *Australasian Journal of Philosophy* 63:465-472.

Weirich, Paul (1986), "Expected Utility and Risk", *British Journal for the Philosophy of Science* 37:419-442.

Wilson, Richard and EAC Crouch (1987), "Risk Assessment and Comparisons: An Introduction", *Science* 236:267-270.

Witte, Eberhard (1972), "Field research on complex decision-making processes - the phase theorem", *International Studies of Management and Organization*, 156-182.

von Wright, Georg Henrik (1963) *The Logic of Preference*.

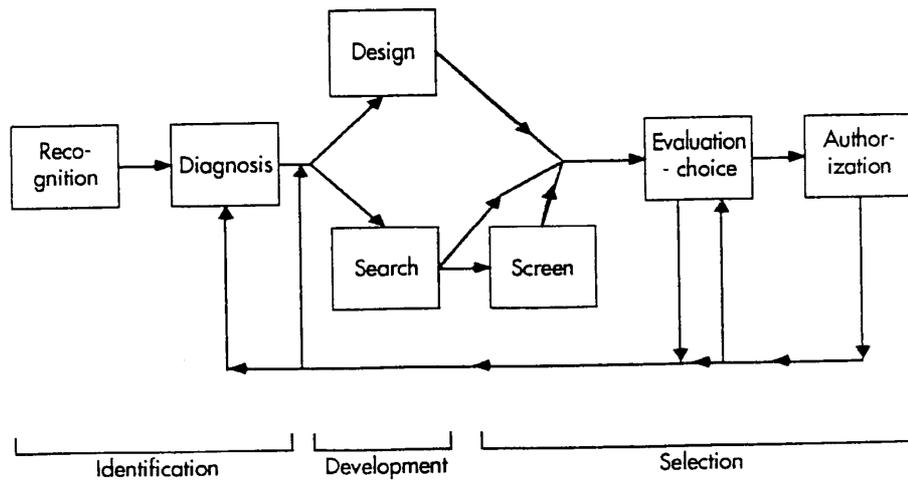


Diagram 1. The relationships between the phases and routines of a decision process, according to Mintzberg et al (1976).

Condorcet	First discussion	Second discussion	Resolution			
Simon	Intelligence	Design	Choice			
Mintzberg et al	Recognition	Diagnosis	Search/Design	Screen	Evaluation-choice	Authorization
Brim et al	Identification	Obtaining information	Production of solutions	Evaluation	Selection	

Diagram 2. A comparison of the stages of the decision process according to Condorcet, Simon, Mintzberg et al and Brim et al.

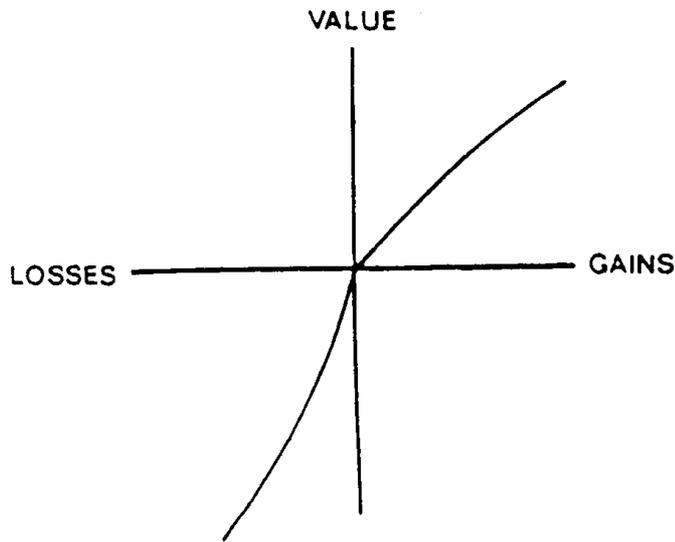


Diagram 3. The value function in prospect theory. (After Kahneman and Tversky [1979] 1988, p. 202.)

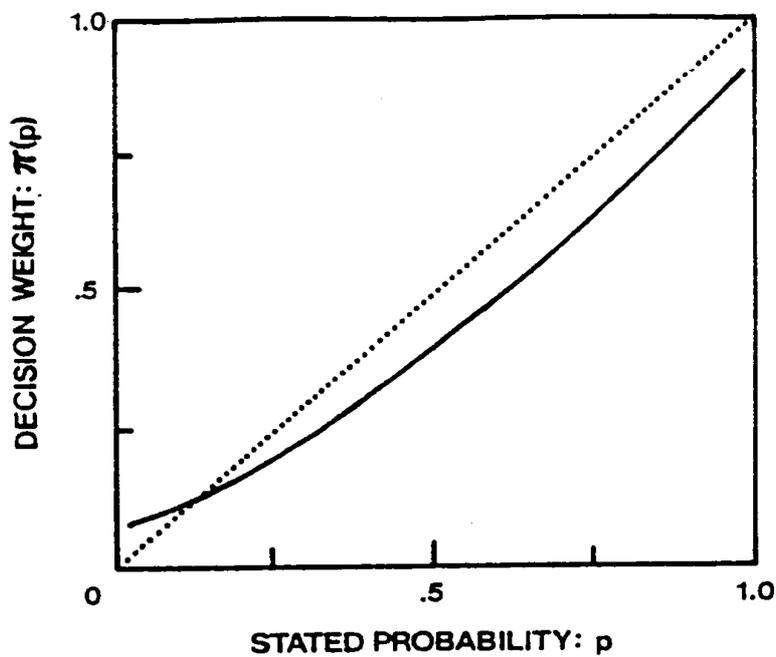


Diagram 4. The decision weight as a function of objective probabilities, according to prospect theory. (After Tversky and Kahneman 1986, p. 264.)

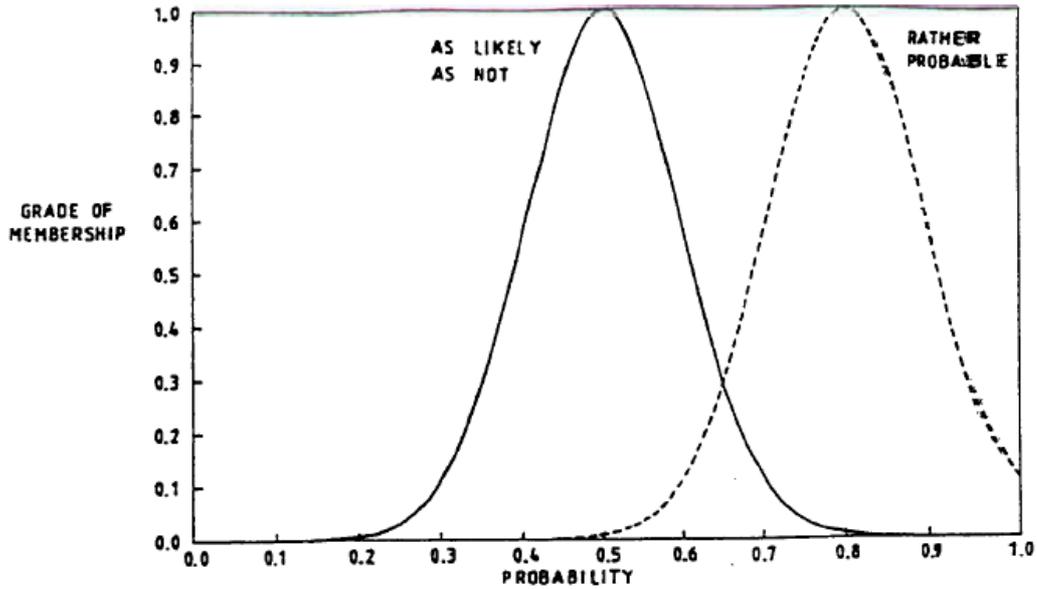


Diagram 5. The vagueness of expert judgments as represented in fuzzy decision theory. (Unwin 1986, p. 30.)

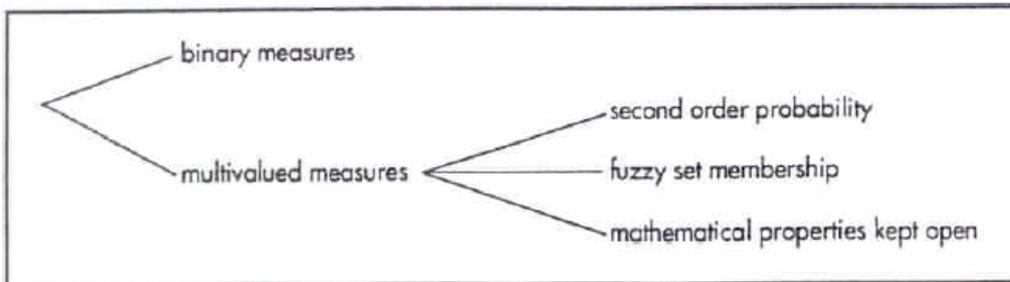


Diagram 6. The major types of measures of incomplete probabilistic information.