

Transformative Decision Rules
Foundations and Applications

Martin Peterson

STOCKHOLM 2003

To Louise and Pamela

ABSTRACT

Peterson, M. 2003. Transformative Decision Rules: Foundations and Applications. *Theses in Philosophy from the Royal Institute of Technology 3*. 72 + x pp. Stockholm. ISBN 91-7283-460-9.

A transformative decision rule alters the representation of a decision problem, either by changing the sets of acts and states taken into consideration, or by modifying the probability or value assignments. Examples of decision rules belonging to this class are the principle of insufficient reason, Isaac Levi's condition of E-admissibility, Luce and Raiffa's merger of states-rule, and the *de minimis* principle. In this doctoral thesis transformative decision rules are analyzed from a foundational point of view, and applied to two decision theoretical problems: (i) How should a rational decision maker model a decision problem in a formal representation ('problem specification', 'formal description')? (ii) What role can transformative decision rules play in the justification of the principle of maximizing expected utility?

The thesis consists of a summary and seven papers. In Papers I and II certain foundational issues concerning transformative decision rules are investigated, and a number of formal properties of this class of rules are proved: convergence, iterativity, and permutability. In Paper III it is argued that there is in general no unique representation of a decision problem that is strictly better than all alternative representations. In Paper IV it is shown that the principle of maximizing expected utility can be decomposed into a sequence of transformative decision rules. A set of axioms is proposed that together justify the principle of maximizing expected utility. It is shown that the suggested axiomatization provides a resolution of Allais' paradox that cannot be obtained by Savage-style, nor by von Neumann and Morgenstern-style axiomatizations. In Paper V the axiomatization from Paper IV is further elaborated, and compared to the axiomatizations proposed by von Neumann and Morgenstern, and Savage. The main results in Paper VI are two impossibility theorems for catastrophe averse decision rules, demonstrating that given a few reasonable desiderata for such rules, there is no rule that can fulfill the proposed desiderata. In Paper VII transformative decision rules are applied to extreme risks, i.e. to a potential outcome of an act for which the probability is low, but whose (negative) value is high.

Key words: transformative decision rule, problem specification, framing, expected utility, decision theory

List of Papers

This doctoral thesis consists of the following summary and the papers:

- I Peterson, M. “Transformative Decision Rules”, *Erkenntnis* 58(1), 71–85, 2003.
- II Peterson, M. “Transformative Decision Rules, Permutability, and Non-Sequential Framing”, *Synthese*, in press, 2003.
- III Peterson, M. “Rival Representations of Decision Problems”, *submitted manuscript*.
- IV Peterson, M. “An Argument for the Principle of Maximizing Expected Utility”, *Theoria* 68(2), 112–128, 2002.
- V Peterson, M. “From Outcomes to Acts: A Non-Standard Axiomatization of the Expected Utility Principle”, *submitted manuscript*.
- VI Peterson, M. “The Limits of Catastrophe Aversion”, *Risk Analysis* 22(3), 527–538, 2002.
- VII Peterson, M. “New Technologies and the Ethics of Extreme Risks”, *Ends and Means* 6(1), 22–30, 2001.

Paper I and Paper II are reprinted with kind permission of *Kluwer Academic Publishers*. Paper IV is reprinted with kind permission of *Theoria*. Paper VI is reprinted with kind permission of *Society for Risk Analysis*. Paper VII is reprinted with kind permission of *Ends and Means*.

Preface

This doctoral thesis is composed of seven papers that deal with different aspects of a particular class of decision rules, *transformative decision rules*, which have seldom been the subject of precise formal analysis. There is also a lengthy summary, placed before the papers, which can be read as a free-standing essay. The official purpose of the summary is to identify the main ideas in the papers and show how they are related. However, I have also used the summary to clarify and discuss in more detail certain points that it has been hard to pay appropriate attention to in the papers, due to the space constraints of academic journals. As a consequence, the focus of the summary is slightly different than that of the papers, as is the structure in which the material is presented.

I wish to express my deep gratitude to my supervisor, Professor Sven Ove Hansson, for spending extensive time reading and discussing this work with me. Without his generous support, and extremely helpful comments and suggestions, I would not have been able to write this thesis.

I also wish to sincerely thank Dr John Cantwell, who has assisted Professor Hansson with valuable advice in the supervision of this thesis. Dr Cantwell's comments have helped me to focus on substantial philosophical issues, and his many corrections have saved me from several embarrassing fallacies.

Thanks are also due to Gustaf Arrhenius, Lars Bergström, Eric Carlsson, Gert Helgesson, Ulrik Kihlbom, Per-Erik Malmnäs, Hans Mathlein, and Per Sandin, for their comments on early versions of

some of the papers. I also wish to thank Åsa Boholm, Jana Fromm, and Lennart Sjöberg for stimulating discussions, and Eleonor Westman for technical assistance.

Certain results reported in this thesis have been presented at a number of symposia and international conferences. I wish to thank the participants for their stimulating and fruitful comments. I also wish to thank the participants of Lars Bergström's seminar at Stockholm University, who have discussed several of the papers with me on a number of occasions.

Finally, I wish to express my sincere thanks to my colleagues in the Philosophy Unit of the Royal Institute of Technology, especially Jonas Clausen, for taking so much interest in my work.

This work has been financially supported by the Bank of Sweden Tercentenary Foundation, through the project *Neglected Risks*. This support is also gratefully acknowledged.

Contents

Abstract	v
List of Papers	vi
Preface	vii
1 Introduction	1
1.1 Transformative and effective decision rules	2
1.2 Two aims	4
1.3 Related research	5
1.4 The structure and scope of this thesis	8
2 Representing Decision Problems	9
2.1 The structure of formal decision problems	10
2.2 Savage's and Jeffrey's set-up	12
2.3 Causal vs. evidential decision theory	13
3 Decision Rules as Functions	17
3.1 Transformative and effective decision rules defined	18
3.2 Composite decision rules	19
3.3 The solution of a decision problem	21
4 Choosing a Representation	23
4.1 A comparison structure for formal representations	23
4.2 Permutability and other structural properties	25
4.3 Four values	29
4.4 Rival representations	31
5 Maximizing Expected Utility	35
5.1 The long run vs. the axiomatic approach	36
5.2 General remarks on axiomatized theories	38
5.3 Summary of previous axiomatizations	39
5.4 Preview of the proposed axiomatization	45
5.5 A complete and unified decision theory	50
6 Comments on Papers I-VII	53
Appendix	61

1 Introduction

In Leonard Savage’s seminal work *The Foundations of Statistics* it is pointed out that rational decision making can be divided into two phases. Expressed in Savage’s terminology, the main objective of the first phase is to decide “which world to use in a given context”¹, that is, to choose an appropriate representation of states, consequences, and acts, and thereby obtain a “formal description, or model, of what the person is uncertain about”.² The purpose of the second phase of rational decision making is to establish “criteria for deciding among possible courses of action”³, and choose an act prescribed by such a criterion. Most decision theorists believe that the principle of maximizing expected utility is the right criterion to opt for.

In the present thesis I introduce and explore a distinction between two classes of decision rules, viz. transformative and effective decision rules, that I use for analyzing the two phases of rational decision making in more detail. The upshot is a decision theory that is *comprehensive* in the sense that it ranges over both phases of rational decision making, and is *unified* in the sense that transformative decision rules are used throughout, thereby making the distinction between the two phases “invisible” from the decision maker’s point of view.

The focus is on foundational issues, i.e. on underlying characteristics of reasonable principles rather than on specific principles. An exception is the principle of maximizing expected utility, for which

¹ Savage 1954/72, p 9.

² *Ibid.*, p 7.

³ *Ibid.*, p 6.

an axiomatic argument based on transformative decision rules is constructed. An advantage of the new axiomatization outlined herein is that it does not employ the notorious ‘sure-thing’ principle or independence axiom.

1.1 Transformative and effective decision rules

As a preliminary definition, a decision rule is *transformative* just in case it alters the representation of a given decision problem by adding, modifying, or deleting information, whereas it is *effective* just in case it singles out a set of recommended acts. In subsequent chapters it will be shown that most effective decision rules comprise transformative elements.

Examples of effective decision rules include the principle of maximizing expected utility, the maximin rule, and the dominance principle.⁴ The following are some well-known examples of decision rules that have a clear transformative structure.

1. Expansion of Alternatives: *If a set of alternative acts is not exhaustive, then it should be expanded by an act not included in the initial set of acts.*
2. The Principle of Insufficient Reason: *If there is no reason to believe that some state in a set of mutually exclusive and exhaustive states is more probable than another, then equal probabilities should be assigned to all states.*
3. Merger of States: *If two or more states yield identical outcomes under all acts in a decision problem under uncertainty, then these repetitious states should be collapsed into one state.*

⁴ These rules are discussed in most textbooks on decision theory, see e.g. Clemen 1991 or Resnik 1993.

4. Levi's Condition of E-Admissibility: *If an alternative act is not E-admissible⁵, then it should be deleted from the set of alternative acts.*
5. De Minimis: *If the probability for some state is sufficiently small, then it should not be included in the set of states considered by the decision maker.*
6. The Precautionary Principle: *If there is a non-zero probability that the outcome of an alternative act is very bad, i.e. below some constant c , then that act should be deleted from the set of alternatives.*

The expansion-of-alternatives-rule is sometimes mentioned *en passant* in textbooks in decision theory.⁶ In many cases though no substantial rationale for this rule is given.

The second example, the principle of insufficient reason, is closely associated with Laplace, who argued that it is a sound principle provided that the decision problem is reasonably described.⁷

The third rule, merger of states, was proposed in a slightly different form in a paper by Milnor in the 1950's, and adopted as Axiom 11 in Luce and Raiffa's discussion of decisions under uncertainty.⁸

Isaac Levi's condition of E-admissibility is the fourth example. It is one of several rules of the same transformative structure proposed by Levi in *The Enterprise of Knowledge*.⁹

The fifth example is the *de minimis* principle. Its philosophical roots can be traced back to Buffon's discussion of 'morally impossible' events in his resolution of the St Petersburg paradox.¹⁰ The phrase '*de minimis*' was derived by risk analysts in the 1970's from the legal

⁵ An act a is, roughly put, E-admissible just in case there is a "seriously permissible" probability function q in B (B is a set of probability functions) and a "seriously permissible" utility function u in G (G is a set of utility functions) such that the expected utility of a is optimal. For a precise definition, see Levi 1980, pp 96.

⁶ Cf. Clemen 1991 and Resnik 1993.

⁷ See Milnor 1954.

⁸ Luce and Raiffa 1957, Chapter 13; Milnor 1954.

⁹ Levi 1980, pp 96.

¹⁰ Arrow 1951, p 414.

principle ‘*de minimis non curat lex*’, which roughly means that the law should not concern itself with trifles.¹¹

The precautionary principle is the last example. The formulation advocated here is one of many alternative formulations.¹² According to this formulation; precaution primarily deals with transforming one representation of a decision problem into another, rather than selecting a particular act.

Modern decision theory has, at least when it comes to foundational issues, been almost exclusively concerned with effective decision rules. This is unfortunate, since transformative decision rules are at least as important as effective ones in choosing which act to perform, because of their capacity to modify the formal representation of a decision problem. For instance, the principle of maximizing expected utility may prescribe a particular set of acts in one representation of a decision problem, but another set of acts in a different representation where the *de minimis* principle has been applied to delete highly improbable states.

1.2 Two aims

The primary aim of this thesis is to analyze transformative decision rules from a foundational point of view. Thereby more clarity will be brought to decision-theoretical issues that in the past have seldom been the subject of precise formal analysis. Some of the questions that will be discussed are: What conditions should a transformative decision rule meet in order to be normatively reasonable? Does it matter in which order the elements of a set of transformative decision rules are applied? To which initial representation of a decision problem should transformative decision rules be applied? How many times should each normatively reasonable transformative decision rule be applied?

¹¹ Whipple (ed.) 1987.

¹² Cf. Sandin 1999.

A secondary aim is to show how transformative decision rules can be applied for analyzing a well-known example of an effective decision rule, namely the principle of maximizing expected utility. The upshot will be an axiomatic argument for the expected utility principle that differs in significant respects from the axiomatizations constructed by e.g. Ramsey, von Neumann and Morgenstern, Savage, and their contemporary successors.¹³ For example, the axiomatization constructed here does not rely on any version of the much criticized independence axiom or sure-thing principle, presupposed in the axiomatizations presented by these authors. Moreover, the new axiomatization is more straightforward than earlier ones in that it does not aim at proving a representation theorem and a corresponding uniqueness theorem (for preferences among lotteries). Instead, it employs the concept of transformative decision rules for decomposing the principle of maximizing expected utility into a sequence of normatively reasonable subrules.

Pursuing the two aims described above results in a decision theory in which choices are made by transforming one representation of a decision problem into another. In such transformations, transformative decision rules themselves become bearers of substantial normative content. In fact, one can maintain that transformative decision rules become bearers of *all* normative content, since only one, extremely uncontroversial, effective decision rule needs to be employed (the **pick** rule defined in Section 3.2).

1.3 Related research

As far as I am aware, transformative decision rules have in the past never been the subject of a precise formal analysis, even though many particular examples of transformative decision rules have been carefully analyzed by other authors.¹⁴ However, the general problems of

¹³ References for most standard axiomatizations of the principle of maximizing expected utility can be found in Chapter 5.

¹⁴ See Section 1.1 for references.

how to model a decision problem in a formal representation and how to justify a particular effective decision rule have been extensively discussed. The earliest explicit discussions of both these issues is found in Savage's *The Foundations of Statistics*. His axiomatization of the principle of maximizing expected utility is the natural point of departure for much modern research in normative decision theory. However, when it comes to the modelling of decision problems in formal representations, it is an overstatement to say that he presented a full-fledged theory about this issue; it is fair to say that his work contains some important observations that are relevant to this problem area. For example, Savage pointed out the importance of restricting representations of decision problems to 'small worlds'—i.e. to isolated situations—since the 'grand' decision problem (How should one live the rest of one's life, from now until death?) is clearly too complex for ordinary decision makers to resolve.

In the wake of Savage's work, countless contributions to normative decision theory have been presented, some of which draw explicitly on ideas put forward by Savage's forerunners Ramsey, De Finetti, and Von Neumann and Morgenstern. Another decision theorist in the same tradition, who has exerted much influence in the past decades, is Richard Jeffrey. He has famously claimed that normative decision theory should be evidential rather than causal, and that formal representations of decision problems should be constructed by using propositions rather than sets of states, consequences, and acts. In Chapter 2 I address these claims in more detail.

Outside the rather limited field of decision theory, defined in a narrow sense, other scholars have done important research on certain basic concepts employed in formal representations of decision problems. For example, in the late 1960's Lars Bergström initiated a debate on the meaning of 'alternative acts'.¹⁵ Unfortunately, as in many other areas of philosophy, that debate has not yet yielded a consensus on the

¹⁵ Bergström 1966.

meaning of this concept.¹⁶ Still other conceptual discussions relevant to the subject of this thesis are those dealing with the meaning of ‘probability’ and ‘utility’, respectively. However, in the present thesis I shall not take a stand on the prevailing debate on objective versus subjective probabilities, and I shall only make a few minor comments on the concept of utility. As already mentioned, the focus herein is on foundational issues concerning transformative decision rules themselves, not on the content put into them, and most of what will be said here is therefore acceptable to adherents of most reasonable views about probability and utility. In particular, I will not presume that a decision maker’s utility function has to be constructed in the way proposed by von Neumann and Morgenstern, Savage, and their successors.¹⁷

Most of the rather sparse literature dealing with the representation of decision problems favours sequential theories.¹⁸ In this context ‘sequential’ means that a proposed process can be divided into a finite number of steps performed in a specified order. A sequential theory may, for example, prescribe that the agent should: 1) Identify the set of alternative acts, and make sure that the acts are mutually exclusive and exhaustive. 2) Identify the relevant set of (act-independent) states of nature, and make sure that the states are mutually exclusive and exhaustive. 3) For each pair of acts and states, identify the corresponding outcome, and assess its utility on an interval scale. 4) Assess the probability of each state and/or outcome.

Even though a sequential theory may seem intuitively natural very few, if any, arguments have been presented in favour of this type of theory. For example, why could one not start by identifying a set of outcomes and thereafter figure out what acts and states would produce those outcomes?

In this thesis a non-sequential theory is defended, in which the order

¹⁶ See Carlsson 1995 for an overview.

¹⁷ See von Neumann and Morgenstern 1947, and Savage 1954/72.

¹⁸ See e.g. Brim *et al* 1962, Resnik 1993 chapter 1.2, and Clemen 1991.

of the different steps (transformative decision rules) is irrelevant.

1.4 The structure and scope of this thesis

This summary, which comprises five more chapters, is followed by seven papers. The latter are referred to by the Roman numerals I–VII, and are commented upon in Chapter 6. In Chapter 2 the distinction among decision problems *per se* and their formal representations is spelled out and explained. In Chapter 3, transformative and effective decision rules are defined in terms of mathematical functions, and in Chapter 4 certain structural constraints on transformative decision rules are proposed and analyzed. In Chapter 5 various axiomatizations of the principle of maximizing expected utility are reviewed.

The scope of this thesis is limited to individual, non-sequential, decision making under risk and uncertainty, i.e. to two of the most elementary types of decision problems. Other areas of decision theory, such as sequential decision making, group decisions, multiattributive decision making, causal decision theory, and theories about bounded rationality, will not be discussed at all or only to a very small extent.

Naturally, not every facet of transformative decision rules will be discussed. For example, as explained in Section 1.3, I shall take no definitive stand on the controversies over the concepts of ‘alternative act’, ‘probability’ and ‘utility’. The focus here is on those issues that adherents of all, or nearly all, positions should be interested in. Throughout I will refrain from extensive discussions of particular transformative decision rules, since readers interested in such investigations can find a wealth of material in the literature.

2 Representing Decision Problems

A decision problem is constituted by the entities of the world that prompt the decision maker to make a choice, as well as those entities that are relevant to each choice. A formal representation of a decision problem, on the other hand, is constituted by a set of symbols representing those entities. Consider, for example, Savage's famous omelet-example, in which the decision maker has just broken five good eggs into a bowl, and has to decide whether or not to break a sixth egg, that is either rotten or not, into the omelet. Depending on the decision maker's choice, and the condition of the sixth egg, the outcome will be a six-egg omelet, a five-egg omelet, or no omelet at all.¹⁹ In this example, the decision problem in question is constituted by a number of concrete real-world objects (the bowl and the six eggs), as well as certain abstract entities (the desire to cook an omelet, and the lack of knowledge about the condition of the sixth egg). A formal representation of the omelet-example is, as explained above, made up of symbols representing the entities included in the decision problem. Of course, there might be more than one way to construct such a formal representation; Savage's suggestion is presented in his 1954/72, p 14. Quite often formal representations of decision problems are visualized in decision matrices or decision trees.

To model one's decision problem in a formal representation of *some* sort is essential to decision theory, since decision rules are only defined relative to such formal representations. It makes no sense to say that

¹⁹ Savage 1954/72, pp 13-15. I have altered his example slightly.

e.g. the principle of maximizing expected utility recommends one act over another, unless there is a formal representation listing the available acts, the possible states of the world, and the corresponding probability and utility functions.

2.1 The structure of formal decision problems

The distinction between decision problems and formal representations of decision problems is commonplace in decision theory, even though it is sometimes expressed in other words. Richard Jeffrey, for instance, uses the term ‘formal Bayesian decision problem[s]’, which he defines as “two rectangular arrays (matrices) of numbers which represent probability and desirability assignments to the act-condition pairs.”²⁰ Savage prefers to speak of ‘formal descriptions’ of decision problems, whereas Michael Resnik uses the term ‘problem specification’. For simplicity, I shall sometimes use the terms ‘formal decision problem’ respectively ‘formal representation’ as synonymous with the somewhat clumsy term ‘formal representation of a decision problem’. All three terms have the same meaning.

In what follows a formal decision problem will be conceived of as an ordered quadruple $\pi = \langle A, S, P, U \rangle$. The intended interpretation of its elements is as follows.

$A = \{a_1, a_2, \dots\}$ a non-empty set of acts.

$S = \{s_1, s_2, \dots\}$ a non-empty set of states.

$P = \{p_1 : A \times S \rightarrow [0, 1], p_2 : A \times S \rightarrow [0, 1], \dots\}$ a set of probability functions.²¹

$U = \{u_1 : A \times S \rightarrow Re, u_2 : A \times S \rightarrow Re, \dots\}$ a set of utility functions.

²⁰ Jeffrey 1983, p 5.

²¹ I take for granted that all elements in P satisfy the axioms of the probability calculus. (Otherwise p_1, p_2, \dots would not, of course, be probability functions.)

A formal decision problem under *risk* is a quadruple $\pi = \langle A, S, P, U \rangle$ in which each of P and U has exactly one element, and a formal decision problem under *uncertainty* is a quadruple $\pi = \langle A, S, P, U \rangle$ in which $P = \emptyset$ and U has exactly one element. Since P and U are sets of functions, rather than single functions, I allow for decision makers to consider several alternative probability and utility measures in a formal decision problem. This set-up can thus model what is sometimes referred to as ‘epistemic risks’, i.e. cases in which there are several alternative probability and utility functions that describe a given situation. (Cf. Levi 1980, Gärdenfors and Sahlin 1982.)

Note that the concept of an ‘outcome’ or ‘consequence’ of an act is not explicitly employed in a formal decision problem. Instead, utilities are assigned to ordered pairs of acts and states. Also note that I have not assumed that the elements in A respectively S (which may be either finite or countable infinite sets) have to be jointly exhaustive and exclusive. To determine whether such requirements ought to be levied or not is a normative issue that will be analyzed more in detail in subsequent chapters.

It is easy to jump to the conclusion that the set-up proposed here presupposes that the probability and utility functions have been derived *ex ante* (i.e. before any preferences among acts have been stated), rather than *ex post* (after the decision maker has stated his or her preferences among the available acts). If correct, this would be a drawback, since the *ex post* approach has been preferred by several influential decision theorists, e.g. Ramsey, von Neumann and Morgenstern, Savage, and Jeffrey. However, the set-up advocated here is, in fact, neutral with respect to the distinction between *ex ante* and *ex post* approaches. The sets P and U are allowed to be empty at the beginning of a representation process, and can be successively expanded with *ex post* functions obtained from preferences among gambles stated by the decision maker at a later stage. This is consistent with what was said in Chapter 1, namely that the present thesis does

not advocate any particular notion of probability or utility, since the focus here is on foundational issues.²²

2.2 Savage's and Jeffrey's set-up

The set-up proposed in the previous section is one of many alternatives. In Leonard Savage's theory, for instance, the fundamental building blocks of a formal decision problem are taken to be a set S of states of the world and a set F of consequences. Acts are then defined as functions from S to F . No probability or utility functions are included in the formal decision problem. They are instead derived "within" the theory by using the decision maker's preferences among hypothetical gambles (that is, acts). The main difference in the set-up chosen by von Neumann and Morgenstern is that they include an exogenously defined probability function in the formal decision problem. The utility function is, however, derived from the premises of the theory itself.

Richard Jeffrey's theory is, in contrast to Savage's (and von Neumann and Morgenstern's), homogenous in the sense that all elements of a formal decision problem—e.g. acts, consequences, probabilities, and utilities—are defined on the same set of entities, namely a set of propositions. For instance, "[a]n act is . . . a proposition which it is within the agent's power to make true if he pleases", and to hold it probable that it will rain tomorrow is "to have a particular attitude toward the proposition that it will rain tomorrow".²³ In line with this, the conjunction $B \wedge C$ of the propositions B and C is interpreted as the set-theoretic intersection of the possible worlds in which B and C are true, and so on.

Note that Jeffrey's definition of an act implies that all consequences of an act performed in a formal decision problem under certainty are acts themselves, since the decision maker can make those propositions

²²An exception is Paper V, in which a particular notion of utility is advocated.

²³ Jeffrey 1983, pp 59, 84.

true if he pleases. Hence, the distinction between acts and consequences cannot be upheld in that case, which appears to be a shortcoming.

Nevertheless, the homogenous character of Jeffrey's set-up is not a compelling reason for preferring it to Savage's, since the latter can easily be reconstructed as a homogenous theory by widening the concept of states such that consequences also fall into this category. The consequence of having a six-egg omelet can, for example, be conceived of as a state in which the agent enjoys a six-egg omelet; thus acts can then be defined as functions from states to states. A similar maneuver can, *mutatis mutandis*, be carried out for the quadruple $\langle A, S, P, U \rangle$.

Arguably, it is more reasonable to take symbols denoting states, instead of full-fledged propositions, to be the basic building blocks of formal decision problems, since states are what ultimately concern decision makers, and states are less opaque from a metaphysical point of view. Since propositions have no spatio-temporal location, as opposed to states (and sentences), we cannot become directly acquainted with them.

2.3 Causal vs. evidential decision theory

The main reason for conceiving of a formal decision problem as a quadruple $\langle A, S, P, U \rangle$, rather than as proposed by Savage or Jeffrey, is that the former set-up is neutral with regard to the controversy between causal and evidential decision theory. In Savage's theory, which has been claimed to be "the leading example of a causal decision theory"²⁴, the probabilities of the states must be independent of the acts, since acts are conceived of as functions from states to consequences. This requirement makes sense only if one holds that decision mak-

²⁴ See Broome 1999, p 103. It can be questioned whether it is correct that Savage's theory is causal. When Savage wrote his book the distinction between causal and evidential decision theory had not yet been drawn, and Savage never explicitly claims that any causal relations are modelled in his theory. Some would prefer to call Savage's theory classic rather than causal.

ers should take beliefs about causal relations into account: acts and states are, in this type of theory, two independent entities that together *cause* outcomes. In an evidential theory such as Jeffrey's, the probability of a certain consequence is allowed to be affected by what act is chosen; the decision maker's beliefs about causal relations play no role. A well-known example in which evidential and causal decision theories recommend different acts is Newcomb's problem.²⁵ Another, more realistic example, is the smoking-caused-by-a-gene-defect problem:²⁶ Suppose that there is some gene defect that is known to cause both lung cancer and smoking. Then the fact that 80% of all smokers suffer from lung cancer should not prevent a causal decision theorist from starting to smoke, since (i) one either has that gene defect or not, and (ii) there is a small enjoyment associated with smoking, and (iii) the probability of lung cancer is not affected by one's choice.²⁷ An evidential decision theorist would, on the contrary, conclude (incorrectly) that if *you* start smoking there is a 80% risk that *you* will contract lung cancer.

It seems obvious that causal decision theory, but not its evidential rival (in its most naïve version), comes to the right conclusion in the smoking-caused-by-gene-defect problem. Some authors have proposed that for precisely this reason evidential decision theory should be interpreted as a theory of valuation rather than as a theory of decision.²⁸ A theory of valuation ranks a set of alternative acts with regard to how good or bad they are in some relevant sense, but it does not prescribe any acts. Valuation should, arguably, be linked to decision in the end, but there is no conceptual mistake involved in separating the two questions.

A significant advantage of Jeffrey's evidential theory, both when interpreted as a theory of decision and as a theory of valuation, is

²⁵ Newcomb's problem was introduced in Nozick 1969.

²⁶ According to Pollock 2002, pp 143-144, this example was first mentioned by Robert Stalnaker in letter to David Lewis in 1978.

²⁷ For an excellent discussion of causal decision theory, see Joyce 1999.

²⁸ See e.g. Broome 1999, pp 104-6.

that one does not have to understand the concept of causality. That concept plays no role in evidential theories.

Since there are significant pros and cons of both the causal and evidential approach, it seems that in a discussion of transformative decision rules one should opt for a set-up that allows for both kinds of theories. Thus, advocates of causal decision theory should claim that the functions $P = \{p_1 : A \times S \rightarrow [0, 1], p_2 : A \times S \rightarrow [0, 1], \dots\}$ ought to be “inert” with respect to A , i.e. that probabilities can equally well be described by a set of functions $P' = \{p'_1 : S \rightarrow [0, 1], p'_2 : S \rightarrow [0, 1], \dots\}$. Evidential decision theorists like Jeffrey should, on the contrary, insist that probabilities indeed have to be represented by functions that take pairs of acts and states as their arguments.

3 Decision Rules as Functions

There is a strong tradition in decision theory to conceive of decision rules as mathematical functions, i.e. as relations that uniquely associates members of one set, *the argument set*, with members of another set, *the value set*. The argument set is usually taken to be a set of formal decision problems, or a set of propositions describing formal decision problems, whereas the value set can be conceived of as a set of alternative acts, or a set of propositions describing acts. Here I accept the general idea that decision rules are mathematical functions, but widen the concept of a decision rule by allowing for rules that are functions from one set of formal decision problems to another such set, i.e. that do not return any set of alternative acts. As indicated in Chapter 1, this novel kind of decision rules will be called transformative decision rules, whereas functions of the former sort will be called effective decision rules. For example, the principle of insufficient reason is a transformative decision rule, i.e. a function, that takes a formal decision problem under uncertainty as its argument and returns the corresponding formal decision problem under risk in which every state is assigned equal probability. The principle of maximizing expected utility is an effective decision rule that takes a formal decision problem as its argument and returns the set of alternative acts having the highest expected utility.

An advantage of conceiving of decision rules as mathematical functions is that it avoids the possibility of a pathological decision rule returning random output. By definition, any function must yield the

same output (i.e. recommended acts or formal decision problems) every time it is applied to a given argument. Furthermore, this approach precludes what in other normative disciplines is sometimes referred to as particularism, i.e. the claim that optimal acts in different situations need not have anything in common besides their optimality.²⁹

3.1 Transformative and effective rules defined

The distinction between transformative and effective decision rules can be spelled out as follows.

Definition 1. *Let Π be a set of formal decision problems. \mathbf{t} is a transformative decision rule in Π if and only if \mathbf{t} is a function such that for all $\pi \in \Pi$, it holds that $\mathbf{t}(\pi) \in \Pi$.*

Definition 2. *Let Π be a set of formal decision problems. \mathbf{e} is an effective decision rule in Π if and only if \mathbf{e} is a function such that for all $\langle A, S, P, U \rangle \in \Pi$ it holds that $\mathbf{e}(\langle A, S, P, U \rangle) \subseteq A$.*

The two classes of decision rules are mutually exclusive, i.e. no decision rule is both transformative and effective. However, in theory, there may be decision rules that are neither transformative nor effective. For instance, one can imagine a function (decision rule) that takes a formal decision problem as its argument and returns ordered pairs of acts and states. Such non-transformative and non-effective “decision rules” will not be discussed here, primarily because there seems to be no plausible decision theoretic interpretation of them.

Note that transformative as well as effective decision rules operate on formal representations of decision problems, not on decision problems. The difference is subtle, but important. For instance, if it can be equally reasonable to model a decision problem in several different formal representations, as I shall argue in Section 4.4, an effective

²⁹For a recent discussion of particularism in ethics, see e.g. Kihlborn 2002.

rule might very well prescribe different alternative acts depending on which formal representation is employed.

3.2 Composite decision rules

A *composite decision rule* is a decision rule that is made up of other decision rules. For example, consider the rule prescribing that in a formal decision problem under uncertainty one should first apply the principle of insufficient reason (**ir**), and thereafter select an act by applying the principle of maximizing expected utility (**eu**). This composite decision rule can be conceived of as a composite function $(\mathbf{ir} \circ \mathbf{eu})(\pi) = \mathbf{eu}(\mathbf{ir}(\pi))$.

More generally speaking, the concept of composite decision rules can be defined as follows.

Definition 3. *Let \mathbf{x}_i be a transformative decision rule and let \mathbf{x}_j be an effective or transformative decision rule. Then $(\mathbf{x}_i \circ \mathbf{x}_j)(\pi) = \mathbf{x}_j(\mathbf{x}_i(\pi))$ is a composite decision rule.*

It follows from Definitions 1 and 2 that composite decision rules are either transformative or effective. More precisely, if $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ are a number of (composite or non-composite) effective or transformative rules, it holds that:

- (i) If $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ are all transformative decision rules, then $\mathbf{x}_1 \circ \mathbf{x}_2 \circ \dots \circ \mathbf{x}_n$ is a transformative decision rule.
- (ii) If $\mathbf{x}_1, \dots, \mathbf{x}_{n-1}$ are transformative decision rules and \mathbf{x}_n is an effective decision rule, then $\mathbf{x}_1 \circ \dots \circ \mathbf{x}_{n-1} \circ \mathbf{x}_n$ is an effective decision rule.
- (iii) If at least one of $\mathbf{x}_1, \dots, \mathbf{x}_{n-1}$ is an effective decision rule, then $\mathbf{x}_1 \circ \dots \circ \mathbf{x}_{n-1} \circ \mathbf{x}_n$ is undefined.

Throughout this thesis I assume that Π is closed.³⁰ As noted in Sections 1.1 and 1.2, a major observation about transformative decision rules is that almost all decision rules that are usually regarded as non-composite effective rules can be decomposed, and shown to have a transformative element. The principle of maximizing expected utility provides an instructive example: Let **weigh** be a transformative decision rule that transforms a formal decision problem under risk into a formal decision problem under certainty with the same alternative set, such that the utility of every outcome in the latter decision problem equals the weighed sum of the utilities and probabilities for the corresponding outcomes in the former, i.e. $\int_{s \in \mathcal{S}} [p(s) \cdot u(a, s)] ds$. Furthermore, let **max** be an effective rule which recommends that the decision maker choose an alternative act associated with the maximum utility in a formal decision problem under certainty. Then, the principle of maximizing expected utility can be trivially reconstructed as a composite decision rule $\mathbf{eu}(\pi) = (\mathbf{weigh} \circ \mathbf{max})(\pi)$.

Many other effective decision rules can be decomposed in analogous ways, e.g. the maximin rule, the minimax regret rule, and Kahneman and Tversky's prospect rule. For example, the maximin rule can be reconstructed as a composite rule (**min** \circ **max**), in which **min** transforms a decision problem under uncertainty into a new decision problem under certainty where one chooses between the "worst-case" scenarios of the original decision problem. The **max** rule is, of course, identical to the effective subrule of the principle of maximizing expected utility, i.e. it recommends the decision maker to choose an alternative act associated with the maximum utility.

The rule **max** can be further decomposed into two subrules, **maxset** and **pick**. The first subrule, **maxset**, is a transformative decision rule that transforms a decision problem under certainty into another decision problem under certainty in which all non-optimal alternatives have been deleted. The effective subrule **pick** thereafter recommends

³⁰ Π is closed (with respect to \mathbf{t}), if and only, if $\pi' = \mathbf{t}(\pi)$ and $\pi \in \Pi$, then $\pi' \in \Pi$.

the decision maker to pick any of the remaining alternatives, which—in case there is more than one—are of course identical with respect to the utility of the certain outcomes.

3.3 The solution of π

The main reason for applying an effective decision rule to a formal representation of a decision problem is to obtain advice about how to act in situations involving risk or uncertainty. This goal is achieved by finding *solutions* of formal decision problems. The solution of a formal decision problem π , to be written as $\Phi(\pi)$, is defined as follows.

Definition 4. *Let $\pi = \langle A, S, P, U \rangle$ be a formal decision problem. Then, $\Phi(\pi)$ is a solution of π if and only if (i) $\Phi(\pi) \subseteq A$ and (ii) every member in $\Phi(\pi)$ is normatively reasonable to perform.*

Definition 4 calls for some comments. First, $\Phi(\pi)$ is not supposed to be known in advance by the decision maker. Typically, the elements of $\Phi(\pi)$ are determined through a normative analysis of the available acts, e.g. by calculating the expected utilities for each act and making an ordinal or cardinal comparison.

Second, I do not want to exclude (at this stage, at least) the possibility that some formal decision problems lack a solution. In such a situation, which resembles the case with moral dilemmas, $\Phi(\pi)$ can be thought of as the empty set, since the empty set is a subset of every set.

The normative predicate ‘normatively reasonable’—which is taken to be a primitive concept in this context—has been chosen instead of e.g. ‘right’ or ‘rational’ or ‘obligatory’ in order to emphasize that more than one act can be included in the solution of a given formal decision problem. It is convenient to call formal decision problems that have identical solutions *solution-equivalent*.

The concept of transformative decision rules suggests a fruitful strategy for finding solutions of formal decision problems, namely to

formulate a set of transformative decision rules that, in some useful sense, simplifies the formal decision problem without affecting its solution. Those rules can then be applied for transforming the original formal decision problem into one in which it is sufficiently easy to determine the solution; typically one should aim at reaching a formal decision problem under certainty. This strategy is, of course, imperfect in the sense that there is no guarantee that there are any transformative decision rules that make the formal decision problem sufficiently simple to resolve. The value of this strategy, therefore, depends solely on its success in actual applications. The decomposition of the principle of maximizing expected utility reported in Paper IV is, it seems, fairly successful.

4 Choosing a representation

By applying transformative decision rules to a formal representation, decision makers can execute choices among alternative representations of a decision problem. It is a truism that not all alternative formal representations are equally reasonable. It is also a truism that whenever several alternative formal representations are available, one should prefer a formal representation that is more reasonable to one that is less reasonable. In this chapter I give more substance to these truisms by analyzing the conditions under which a sequence of transformative decision rules may be applied to an initial representation of a decision problem, and what makes one formal representation more reasonable than another.

Of course, transformative decision rules are not the only means decision makers can use for choosing among alternative formal representations. But transformative decision rules provide a model for choices among formal representations that is more precise than other models. For example, structural constraints on transformative decision rules can be rendered much more precisely than general rules of thumb or procedures based on direct intuition.

4.1 A comparison structure for formal representations

It is essential to separate those transformative decision rules that may be applied to a formal representation, from those that may not be applied. In order to accomplish this, let $\langle \Pi, \succeq \rangle$ be a comparison

structure for formal decision problems, in which Π is a set of formal decision problems, and \succeq is a relation in Π corresponding to the English phrase ‘at least as reasonable representation as’.³¹ All elements in Π are different formal representations of one and the same decision problem, and \succeq orders the elements in that set with regard to some list of relevant decision theoretic values, e.g. realizability, completeness, relevance and simplicity. (These values are discussed at length in Section 4.3.) I assume that \succeq is reflexive and transitive. Of course, the relations \succ and \sim can be constructed in terms of \succeq in the usual way. Note that in case Π is an infinite set it need not contain an optimal element, i.e. a formal decision problem that is at least as reasonable as all alternative representations.

Arguably, a reasonable initial formal representation, to which the decision maker may apply a suitable sequence of transformative decision rules, is the formal representation containing as many acts, states, probability and utility functions as possible, since an improved formal representation can then be obtained by simply filtering out insignificant or inconsistent information in the subsequent transformative phase.

There are several structural conditions that normatively reasonable transformative decision rules ought to fulfill. Let \mathbf{T} be a set of transformative decision rules and let Π be a set of formal decision problems, such that for every $\mathbf{t} \in \mathbf{T}$ and every $\pi \in \Pi$ it holds that $\mathbf{t}(\pi) \in \Pi$. Then, the most basic normative condition is that for every $\mathbf{t} \in \mathbf{T}$ and every $\pi \in \Pi$ it should hold that:

Condition 1. $\mathbf{t}(\pi) \succeq \pi$

According to Condition 1, it is sufficient that the application of a transformative decision rule yields a formal representation that is at least as good as the one to which it was applied. As an alternative to Condition 1, one could demand that the application of a transformative decision rule should yield a formal decision problem that is

³¹ The concept of comparison structures is investigated in detail in Hansson 2001, Chapter 2.

strictly better than the original one in case the original representation is altered. Formally expressed, this amounts to saying that it ought to hold that $\mathbf{t}(\pi) \neq \pi \Rightarrow \mathbf{t}(\pi) \succ \pi$ for every \mathbf{t} and $\pi \in \Pi$. The reason for requiring strict rather than weak preference in this modified version of Condition 1 is that decision makers ought not to relinquish a formal representation of a decision problem unless there is at least some advantage in the new representation. There is always a risk of making mistakes when a formal representation is transformed into another. On the other hand, the original version of Condition 1 is more reasonable in case several transformative decision rules $\mathbf{t}, \mathbf{u}, \dots$ are available. For instance, it might be the case that \mathbf{u} yields a strictly better formal decision problem if applied to $\mathbf{t}(\pi)$ but not if applied to π , even though $\mathbf{t}(\pi) \sim \pi$. In this case, the requirement that $\mathbf{t}(\pi) \succeq \pi$ is more reasonable than $\mathbf{t}(\pi) \neq \pi \Rightarrow \mathbf{t}(\pi) \succ \pi$, since only the former formulation of Condition 1 allows the decision maker to carry out the transformation from π to $(\mathbf{t} \circ \mathbf{u})(\pi)$. The original version of Condition 1, based on weak preference, is advocated in Paper II, whereas the modified version based on strict preference is adopted in Paper I.

One might reasonably ask whether there are cases in which Condition 1, even when based on weak preference, is unreasonably strong? Suppose, for instance, that $\pi \succ \mathbf{t}(\pi)$ and that $(\mathbf{t} \circ \mathbf{u})(\pi) \succ \pi$. As stated here, Condition 1 does not permit the decision maker to carry out the transformation from π to $(\mathbf{t} \circ \mathbf{u})(\pi)$ in two separate steps. However, note that this condition does not prevent the decision maker from treating the composite rule $(\mathbf{t} \circ \mathbf{u})$ as a single rule fulfilling this condition. Therefore, this is not a counter-example to Condition 1.

4.2 Permutability and other structural properties

Loosely speaking, a set of transformative decision rules is permutable just in case it does not matter in which order the rules are applied. The concept of permutability is interesting primarily because it can signif-

icantly reduce the complexity of the decision maker's choice among formal representations: If the rules are permutable it will be sufficient to determine *which* transformative decision rules the decision maker should apply, whereas the order in which they are applied can be disregarded.

There are two fundamentally different notions of permutability, viz. weak and strong permutability. Weak permutability holds whenever the aggregated value of all formal representations obtained from the different permutations of the rules are equal. Strong permutability holds just in case the order of the rules is irrelevant in the sense that all formal decision problems returned by different sequences of some set of transformative decision rules are sufficiently similar for yielding identical recommendations by a given effective rule. Consider the following formal definitions.

Definition 5. *A set \mathbf{T} of transformative decision rules is weakly permutable for Π just in case, for every $\mathbf{t}, \mathbf{u} \in \mathbf{T}$ and every $\pi \in \Pi$, it holds that $(\mathbf{t} \circ \mathbf{u})(\pi) \sim (\mathbf{u} \circ \mathbf{t})(\pi)$.*

Definition 6. *A set \mathbf{T} of transformative decision rules is strongly permutable for Π with respect to an effective rule \mathbf{e} just in case, for all $\mathbf{t}, \mathbf{u} \in \mathbf{T}$ and every $\pi \in \Pi$, it holds that $(\mathbf{t} \circ \mathbf{u} \circ \mathbf{e})(\pi) = (\mathbf{u} \circ \mathbf{t} \circ \mathbf{e})(\pi)$.*

In what follows I focus on weak permutability, which is the most interesting concept from a general decision theoretic point of view. For a discussion of strong permutability, see Paper II, Section 4.

Below are two rather weak structural conditions that are fulfilled by many transformative decision rules.

Condition 2. $\mathbf{t}(\pi) \succeq \mathbf{t}(\pi') \Leftrightarrow \pi \succeq \pi'$

Condition 3. $\mathbf{t}(\pi) \sim (\mathbf{t} \circ \mathbf{t})(\pi)$

Condition 2 is a monotonicity condition. It implies that transformative decision rules should preserve the order between the elements of a set of formal decision problems. That is, in case a given transformative

decision rule is applied to different formal representations, then the decision maker's preference among those formal representations should not be altered. For an example, see Paper II, Section 3.

Condition 3 states that the aggregated value of a formal decision problem should remain constant no matter how many times (≥ 1) a transformative decision rule is iterated. A detailed defense of this structural condition (formulated in a slightly different way) can be found in Paper I.

Taken together Condition 1–3 implies that every set of transformative decision rules satisfying these conditions are weakly permutable. The following theorem covers the case with two transformative decision rules. The general case with three or more transformative decision rules is dealt with in Paper II, Sect. 3.

Theorem 1. *Let \mathbf{T} be a set of transformative rules for Π that satisfy Conditions 1–3. Then, for every $\mathbf{u}, \mathbf{t} \in \mathbf{T}$ and every $\pi \in \Pi$, it holds that $(\mathbf{u} \circ \mathbf{t})(\pi) \sim (\mathbf{t} \circ \mathbf{u})(\pi)$.*

The proof of Theorem 1 and all other theorems in this summary can be found in the appendix.

Conditions 1–3 are, as stated in Theorem 1, sufficient but not necessary conditions for weak permutability. Consequently, there might be sets of transformative decision rules that are weakly permutable, even though they do not satisfy those three conditions. It is an open question whether there is any set of conditions that is both necessary and sufficient, and that is also reasonably attractive in terms of practical applicability. Note, for example, that not even Condition 1 is a necessary condition for weak permutability.

It is not enough for decision makers to know that the order in which different transformative decision rules are applied is insignificant. For obvious reasons, decision makers need also know how many times each $\mathbf{t} \in \mathbf{T}$ should be applied, and exactly how many of the elements in \mathbf{T} one should apply. In addition to the assertions about weak permutability, Conditions 1–3 also imply that each $\mathbf{t} \in \mathbf{T}$ only needs to

be applied once, and that as many elements as possible in \mathbf{T} can be applied.

In Theorem 2 below, \mathbf{u} can be conceived of as any (composite or non-composite) transformative decision rule satisfying the conditions of the theorem. Hence, nothing can be gained by applying \mathbf{t} more than once, no matter what transformations was carried out by \mathbf{u} .

Theorem 2. *Let \mathbf{T} be a set of transformative rules for Π that satisfy Conditions 1–3. Then, for every \mathbf{t} , $\mathbf{u} \in \mathbf{T}$ and every $\pi \in \Pi$, it holds that $(\mathbf{t} \circ \mathbf{u} \circ \mathbf{t})(\pi) \sim (\mathbf{u} \circ \mathbf{t})(\pi)$.*

Theorem 3 below shows that all permutations obtained from the *largest* subset of rules satisfying Condition 1–3 are optimal. Hence, the decision maker may safely apply *all* transformative decision rules that satisfy these conditions. The following observations are instrumental in the proof of Theorem 3.

Observation 1. *Let \mathbf{T} be a set of transformative rules for Π that satisfy Conditions 1–3. Then, for all \mathbf{t} , $\mathbf{u} \in \mathbf{T}$ and all $\pi \in \Pi$, it holds that:*

1. $(\mathbf{t} \circ \mathbf{u})(\pi) \succeq \pi$
2. $(\mathbf{t} \circ \mathbf{u})(\pi) \succeq (\mathbf{t} \circ \mathbf{u})(\pi') \Leftrightarrow \pi \succeq \pi'$
3. $(\mathbf{t} \circ \mathbf{u})(\pi) \sim (\mathbf{t} \circ \mathbf{u} \circ \mathbf{t} \circ \mathbf{u})(\pi)$

Observation 2. *Let \mathbf{T} be a (finite) set of transformative rules for Π that satisfy Conditions 1–3. Then, for every subset \mathbf{T}' of \mathbf{T} and every $\pi \in \Pi$ it holds that all permutations p_a and p_b obtainable from \mathbf{T}' are of equal value, i.e. $p_a(\pi) \sim p_b(\pi)$.*

Theorem 3. *Let \mathbf{T} be a (finite) set of transformative rules for Π that satisfy Conditions 1–3, and let $\overline{\mathbf{T}}$ be the power set of \mathbf{T} . Then, for all subsets \mathbf{A}, \mathbf{B} of $\overline{\mathbf{T}}$ it holds that if $\mathbf{A} \subseteq \mathbf{B}$ then, for every $\pi \in \Pi$ and every permutation p_a obtainable from \mathbf{A} and every permutation p_b obtainable from \mathbf{B} , it holds that $p_b(\pi) \succeq p_a(\pi)$.*

To sum up, the theorems of this section show that if a set of transformative decision rules satisfy three structural conditions, the decision maker may apply all transformative decision rules in this set, in any desired order he or she wishes, with no requirement to apply any rule more than once.

4.3 Four values

At this point it is instrumental to say a bit more about the factors that determine whether the relation \succeq holds between two alternative formal representations. In Section 4.1 it was proposed that there are at least four values that should be taken into account here, namely: *realizability*, *completeness*, *relevance*, and *simplicity*.³² Note that all these values are epistemic values rather than moral values.³³

A formal decision problem is *realizable* just in case all elements in the representation correspond to features of the decision problem that have a potential to realize. A realizable formal representation should, for example, not include acts that cannot be performed by the decision maker, e.g. the act ‘run 100 meters in five seconds’. This constraint also holds for states of the world: all states listed in a formal representation should of course be possible states of nature, and similar points can be made about probability and utility functions. Depending on how close the formal representation is to what can actually be realized, one can speak of degrees of realizability. Suppose, for instance, that you have decided to put a pencil mark ten centimeters below the line at the top of this page. No matter how hard you try, you will never be able to put the mark exactly ten centimeters below the line. In spite of this, (a representation including) the act of putting a mark ten centimeters below the line has a high degree of realizability, since

³² In Paper I, I only mention two values, viz. accuracy and simplicity, whereas the four values mentioned here are listed in Paper II and III. I now think that the list of four values provides the best account of the significant aspects of a formal representation.

³³ The distinction between moral and epistemic values has been extensively analyzed by Hempel. See e.g. Hempel 2000.

that act is very close to what can actually be realized.

The more *complete* a formal representation is, the fewer (realizable) acts, states, probability or utility functions have been left out. Of course, a formal representation can have a high degree of realizability without also having a high degree of completeness and *vice versa*. Consider for example a formal representation containing the act ‘bring the umbrella’ and the state ‘it rains’. This formal representation is fully realizable, but it has a low degree of completeness.

Relevance means that a reasonable formal representation ought to be faithful to the decision problem under consideration. If, for example, you are about to go for a walk or not go for a walk, and list the acts ‘move my right foot’ and ‘move my left foot’, and add some appropriate states, and probability and utility functions, you may perhaps end up with a realizable and complete formal representation. But it does not fulfill the condition of relevance to a high degree. A more relevant formal representation should rather contain the acts ‘go for a walk’ and ‘stay home’.

The fourth value that ought to be taken into consideration when a decision problem is to be modelled in a formal representation is *simplicity*. A formal representation containing a small number of acts and states is, *ceteris paribus*, simpler than one containing many acts and states. The reason for including simplicity in the list of values is that in case too many features of a decision problem are modelled in a formal representation, it might very well turn out that the calculations and investigations needed for coming to a decision will demand more effort than what is actually motivated by the decision problem. Suppose, for example, that you are about to decide whether you should bring the umbrella or not when going for a walk in the park. It might perhaps be true that a realizable, complete, and relevant formal representation of this decision problem should include the state in which a bird decides to discharge some excrement on the spot where you are, and that an umbrella would protect you from this. However, in case

the number of birds in the park is not extremely high, it is reasonable to neglect this state, for reasons of simplicity.

In some cases the four epistemic values pull in different directions. For example, one formal representation may score high in realizability and simplicity but low in completeness and relevance, while the reverse holds for another representation. In situations of this kind the decision maker has to make a trade-off, and select a formal representation in which he or she considers the aggregated value to be optimal. In this trade-off the decision maker may assign different weights to different values.

4.4 Rival Representations

Certain structural properties of transformative decision rules, especially the concept of weak permutability, gives rise to a rather general problem for normative decision theory. This problem is discussed in Paper III in terms of ‘rival representations’. In order to illustrate the phenomenon of rival representations, suppose that two different formal representations have been obtained by applying two different permutations of a set of weakly permutable transformative decision rules. Then, as described in the previous section, it might be the case that the first formal representation scores high in e.g. realizability and simplicity but low in completeness and relevance, while the reverse holds for the second formal representation. The aggregated value of both formal representations may, however, be equal—and even optimal. That is, the aggregated values of the two different representations might be as high as can possibly be achieved by any formal representation. Furthermore, according to some preferred effective decision rule, e.g. the principle of maximizing expected utility, a given act may be judged as rational (normatively reasonable) by this principle when evaluated in the first formal representation, but judged as non-rational when evaluated by the same principle in the second representation. Theo-

retically, it might even be the case that *all* acts that are rational in one formal representation are non-rational in another formal representation, while *all* acts that are rational in the latter are non-rational in the former. At this point some of us may feel the ground tottering beneath our feet. What should a rational decision maker do in a situation like this? How should one *act*?

To some degree the present problem resembles the problem of underdetermination in science, famously discussed by e.g. W.V.O. Quine.³⁴ According to Quine's notion of underdetermination there might be several different scientific theories that explain all accumulated evidence equally well. In such cases there are at least two possible positions one could take. *Ecumenists* think that both (all) of the incompatible theories should be regarded as 'locally' true, whereas *sectarians* argue that every rivalling theory ought to be considered false. The main problem with the latter standpoint is that it forces us to make judgments about the truth or falsity of scientific theories that are not based on empirical evidence (or other rational considerations, e.g. simplicity, scope, etc.), since the accumulated evidence and all other relevant features of both theories are equal. The problem faced by advocates of the ecumenical position is to explain what it means for incompatible theories to be 'locally' true; in that case truth as "correspondence to external facts" seems impossible.

In decision theory the ecumenical position is less problematic. In fact, it seems perfectly reasonable to maintain that in case one and the same act is judged as rational in one formal representation but non-rational in another formal representation (by the same effective rule), then it is both normatively reasonable ('right', 'correct', etc.) and not normatively reasonable to perform that act. This is not contradictory, given that acts are normatively reasonable only *relative* to a certain formal decision problem.

From an action guiding perspective one can maintain, in a truly

³⁴ See e.g. Quine 1992.

ecumenical vein, that in case an act is judged as both normatively reasonable and unreasonable (in different optimal formal representations) then it is both permitted to perform the act and permitted not to perform the act in question. Only acts that are not normatively reasonable in any optimal formal representation are forbidden.

Arguably, at least some of the disagreement about rational action found in society (e.g. about environmental and health issues) can be explained by the phenomena of rival representations. For instance, some people argue that certain actions should be taken (e.g. build more nuclear power plants) because that maximizes expected utility, whereas others come to the opposite conclusion as they apply the same effective decision rule. In case two different optimal formal representations were used in the calculations, these opposing recommendations would not be hard to explain, nor to cope with: the two alternative strategies are best according to different equally reasonable and optimal formal representations, and therefore any of the strategies may be performed.

5 Maximizing Expected Utility

The purpose of the second phase of rational decision making is to establish criteria for choosing among alternative acts, i.e. to select an effective decision rule and choose an act prescribed by that effective decision rule. Many decision theorists intuitively feel, and some have constructed explicit arguments intended to show, that the principle of maximizing expected utility is the right effective rule to opt for.

A major point in this thesis is that the concept of transformative decision rules allows the decision maker to decompose the principle of maximizing expected utility into a sequence of transformative decision rules followed by a single effective decision rule (the **max** rule formulated in Section 3.2), in which each subrule can be justified individually. In Papers IV and V this procedure is used for constructing a non-standard axiomatization of the principle of maximizing expected utility.

The first scholar to clearly formulate the principle of maximizing expected utility was Pascal, who proposed that it could be applied as a method for calculating the fair price of gambles.³⁵ In the Port-Royal Logic from 1662 it was generalized by Arnoult and Niicole to cover all kinds of decisions:

in order to decide what we ought to do to obtain some good or avoid some harm, it is necessary to consider not only the good or harm itself, but also the probability that it will or will not occur; and to view geometrically the

³⁵ Another early proponent of this decision rule was Bernoulli 1738, who explicitly discussed the concept of expected *utility*, in contrast to expected monetary value.

proportion that all these things have when taken together. . . . Those who do not draw [this conclusion], however exact they are in everything else, are treated in Scripture as foolish and senseless persons, and they misuse logic, reason, and life.³⁶

Today many major decisions by government agencies and large private corporations are taken on the basis of expected utility calculations. For instance, such calculations are routinely used in cost benefit analysis (CBA) for weighing risky benefits against risky costs, as well as in the insurance industry for certain areas of decision making. In order to emphasize the semi-formal nature of the principle of maximizing expected utility, I shall in what follows take this principle to mean that decision makers should choose an act a_i for which $\sum_{j=1}^n p(s_j) \cdot u(a_i, s_j)$ is maximal, where $p(s_j)$ is the probability of state s_j and $u(a_i, s_j)$ the utility of outcome $\langle a_i, s_j \rangle$.

Of course, there is no guarantee that the sequence of transformative decision rules applied in the first phase of rational decision making always returns a formal decision problem under risk. For some decision problems the optimal representation (or *all* optimal representations) might be a formal decision problem under uncertainty. Naturally, the axiomatization constructed here is only applicable when a formal decision problem under risk has been obtained in the first phase.

5.1 The long run vs. the axiomatic approach

Decision theorists have proposed roughly two kinds of justifications for the principle of maximizing expected utility. The point of departure in the classic argument is the law of large numbers. This mathematical theorem states that if a random experiment is repeated n times, and each experiment has the probability p for success, then the probability that the percentage of successes differs from the probability p by more than a fixed positive amount, $\varepsilon > 0$, converges to zero as the number

³⁶Arnauld and Nicole 1662/1996, pp 273-75.

of trials n goes to infinity, for every positive ε . Therefore, the decision maker will be better off in the long run if he chooses to maximize expected utility, rather than opts for any other alternative. A widely recognized weakness of the classic argument is that decision makers seldom face similar decision problems a large number of times. Indeed, it might be reasonably claimed that most decision problems are unique in at least some respect. Consequently, arguments based on what would happen in the long run are commonly regarded to be of little practical relevance. Already Keynes stressed this important point, by observing that “in the long run we are all dead”.³⁷

Modern arguments for the principle of maximizing utility do not rely on the law of large numbers. Instead, the basic idea is to show that the principle of maximizing expected utility can be derived from a number of normatively reasonable axioms, that hold independently of what would happen in the long run. The pioneers in axiomatic utility theory were von Neumann and Morgenstern (1947) and Savage (1954/72), who famously formulated a number of axioms from which they derived representation and uniqueness theorems in support of the principle of maximizing expected utility.³⁸ Later on similar axiomatizations have been constructed by e.g. Marschack (1950), Herstein and Milnor (1953), Davidson et al (1957), Fishburn (1970/79), Harsanyi (1978, 1979), Jeffrey and Bolker (Jeffrey 1983), Kreps (1988), Segal (1990), Oddie and Milne (1991), Schmidt (1998), and Joyce (1999).

³⁷ Keynes 1923, p 89. In this quote Keynes is mainly concerned with long run effects in economic theory. Similar remarks about the irrelevance of such effects in general decision theory can be found in Keynes 1921, Chapter 26.

³⁸ An important forerunner to Savage is Ramsey 1926, who presented an incomplete outline of a similar axiomatic argument almost three decades earlier. However, Ramsey refrained from finishing his argument because “this would, I think, be rather like working out to seven places of decimals a result only valid to two.” (Ramsey 1926, p 180). For a modern presentation of Ramsey’s argument, see Sahlin 1990.

5.2 General remarks on axiomatized theories

In an axiomatized theory a small number of intuitively reasonable propositions are used for justifying some more complex propositions. Two standard examples are Euclid’s axiomatization of geometry and Peano’s axiomatization of arithmetics. In past times it was commonly believed that a successful axiomatization ought to adopt as its axioms some set of *fundamental* and *indisputable* truths. Such very strong demands are seldom stated nowadays, mainly because there appears to be no fundamental and indisputable truths. It seems more reasonable to maintain that even axioms should be regarded as fallible—as every other proposition—and hence compatible with the Quinean slogan that “no statement is immune to revision”.³⁹

According to the suggested view, an axiomatization of a theory is valuable mainly because it shows how certain central ideas are connected to each other. If a set of axioms accepted by an agent is less controversial than some theorem in question, then a proof that the theorem follows from the axioms indicates that the agent should either give up at least one of the axioms, or accept the theorem. In case the axioms are sufficiently reasonable, it will be more attractive to accept the theorem compared to giving up an axiom. Note that this way of conceiving of axiomatizations is fully compatible with epistemological coherentism, according to which agents are justified to believe in (a set of) propositions in proportion to how well the proposition(s) cohere with the rest of the agent’s beliefs.

In decision theory, axiomatized theories frequently take the form of deriving a representation theorem and a corresponding uniqueness theorem. Briefly put, a representation theorem shows that a certain non-numerical structure, e.g. someone’s preferences among a set of risky acts $\{a, b, \dots\}$, can be represented by some real-valued function R , such that $a \succ b \Leftrightarrow R(a) > R(b)$. A uniqueness theorem tells us

³⁹ Quine 1951, Section VI.

what transformations of R are allowed, given that the new function R' is to be a accurate representation of the same structure.

A major point in this thesis is that an axiomatized decision theory does not necessarily has to take the form of deriving a representation theorem and a corresponding uniqueness theorem (for preferences among risky alternatives). On the other hand, it will be argued that it might be at least as fruitful to construct axiomatizations in which the derived theorem(s) have a more direct normative relevance.

Logicians interested in axiomatized theories often focus on two formal properties such systems may have, viz. soundness and completeness. In axiomatic decision theory, however, soundness and completeness are seldom explicitly discussed. Most decision theorists tend to focus their discussions on more down-to-earth matters, such as the intuitive plausibility of some given axiom. However, to some degree, a representation theorem and its corresponding uniqueness theorem can play the same role as a soundness and completeness theorem.

5.3 Summary of previous axiomatizations

In order to see how different axiomatizations of the principle of maximizing expected utility relate to each other, it is fruitful to take a careful look at the systems proposed by von Neumann and Morgenstern (1947) respectively Savage (1954/72).

Von Neumann and Morgenstern

Let Z be a finite set of prizes and let P be the set of functions $p : Z \rightarrow [0, 1]$ such that $\sum_z p(z) = 1$. For every number $\alpha \in (0, 1)$, there is an operation $\alpha p + (1 - \alpha)q$ such that if p and q are both in P , then $\alpha p + (1 - \alpha)q$ is also in P . Intuitively, $P = \{p, q, r, \dots\}$ can be conceived of as the set of all possible lotteries between the elements of Z . Von Neumann and Morgenstern “insist upon the . . . interpretation of probability as frequency in long runs”, and explicitly assume that

the numerical probabilities for all elements of P are known to the decision maker.⁴⁰ In this set-up, an ‘alternative act’ should be thought of as an element in P , i.e. a lottery.

Von Neumann and Morgenstern propose that the following axioms hold for the relation \succ on P :^{41,42}

vNM 1. *For all $p, q \in P$: Either $p \succ q$ or $p \sim q$ or $q \succ p$.*

vNM 2. *For all $p, q \in P$: If $p \succ q$ and $q \succ r$, then $p \succ r$.*

vNM 3. *For all $p, q, r \in P$ and $\alpha \in (0, 1]$: If $p \succ q$ then $\alpha p + (1 - \alpha)r \succ \alpha q + (1 - \alpha)r$.*

vNM 4. *For all $p, q, r \in P$: If $p \succ q \succ r$ then there exist $\alpha, \beta \in (0, 1)$ such that $\alpha p + (1 - \alpha)r \succ q \succ \beta p + (1 - \beta)r$.*

From this set of axioms one can derive the following representation and uniqueness theorem.

Theorem 4. *There is a function u from P to $[0, 1]$ such that:*

(1) *If $p \succ q$, then $u(p) > u(q)$.*

(2) *$u(\alpha p + (1 - \alpha)q) = \alpha u(p) + (1 - \alpha)u(q)$.*

(3) *For every other function u' satisfying (1) and (2), there are numbers $c > 0$ and d such that $u' = c \cdot u + d$.*

⁴⁰ Von Neumann and Morgenstern 1947, pp 19-20.

⁴¹ Actually, von Neumann and Morgenstern expressed their axioms in a slightly different way, see their 1947, pp 24-27. The present formulations, which are more attractive from a technical point of view, can be found in e.g. Kreps 1988, pp 43-44, and Schmidt 1998, pp 4-6.

⁴² Of course, $p \sim q$ means $\neg(p \succ q) \wedge \neg(q \succ p)$.

Property (1) states that the utility function u assigns higher utility numbers to better lotteries. Property (2) is the expected utility property, according to which the value of a compound lottery is equal to the expected value of its components. Property (3) implies that all utility functions satisfying (1) and (2) are positive linear transformations of each other, i.e. that utility is measured on an interval scale.

In the present exposition, the much-criticized independence axiom is expressed by **vNM3**. Numerous examples have been constructed purporting to show that **vNM3** is an unacceptable restriction on preferences among risky acts.⁴³ In addition to this, some authors have also questioned the plausibility of **vNM1** and **vNM4**, see e.g. Schmidt (1998).

It should be emphasized that von Neumann and Morgenstern (1947) expressed no explicit normative intentions when discussing the principle of maximizing expected utility. For them, the primary concern was to give increased credibility to the use of numerical utility functions in game theory. This was achieved by showing that decision makers who satisfy **vNM 1 – vNM 4** can be described *as if* they were making decision by calculating expected utilities. However, in the past decades it has been popular to also open up the door for normative interpretations of their work. See e.g. Kreps (1988), Resnik (1993), Schmidt (1998), Ekenberg et al (2001).

Savage

Savage (1954) is the first author who explicitly claims that his axioms should be given a normative interpretation. A major technical difference compared to von Neumann and Morgenstern (1947) is that Savage makes no use of objective probabilities. Instead, he shows that preferences among risky alternatives can be used for constructing a subjective probability function and a utility function simultaneously.

Let S be a set of states s, s', \dots with subsets A, B, \dots , and let X

⁴³ Cf. Allais 1953, Rabinowicz 1997, Schmidt 1998, pp 15-19.

be a set of consequences x, x', \dots . Acts are conceived of as functions $\mathbf{f}, \mathbf{g}, \dots$ from S to X . It is assumed that \succeq is a relation between pairs of acts. In this framework, \mathbf{f} and \mathbf{g} *agree* with each other in B just in case $\mathbf{f}(s) = \mathbf{g}(s)$ for all $s \in B$. Furthermore, $\mathbf{f} \succeq \mathbf{g}$ *given* B , if and only if, if it were known that B does not obtain, then \mathbf{f} is weakly preferred to \mathbf{g} ; and B is *null*, if and only if, $\mathbf{f} \succeq \mathbf{g}$ given B for every \mathbf{f}, \mathbf{g} . Finally, A is *not more probable* that B (abbreviated $A \leq B$) if and only if $\mathbf{f}_A \succeq \mathbf{f}_B$ or $x \succeq x'$, for every $\mathbf{f}_A, \mathbf{f}_B, x, x'$ such that: $\mathbf{f}_A(s) = x$ for $s \in A$, $\mathbf{f}_A(s) = x'$ for $s \in \neg A$, $\mathbf{f}_B(s) = x$, for $s \in B$, $\mathbf{f}_B(s) = x'$, for $s \in \neg B$.

SV 1. \succeq is a weak order.

SV 2. If \mathbf{f}, \mathbf{g} , and \mathbf{f}', \mathbf{g}' are such that:

- (1) in $\neg B$, \mathbf{f} agrees with \mathbf{g} , and \mathbf{f}' agrees with \mathbf{g}' ,
- (2) in B , \mathbf{f} agrees with \mathbf{f}' , and \mathbf{g} agrees with \mathbf{g}' ,
- (3) $\mathbf{f} \succeq \mathbf{g}$;

then $\mathbf{f}' \succeq \mathbf{g}'$

SV 3. If $\mathbf{f}(s) = x, \mathbf{f}'(s) = x'$ for every $s \in B$, and B is not null, then $\mathbf{f} \succeq \mathbf{f}'$ given B , if and only if, $x \succeq x'$.

SV 4. For every A and B : $A \geq B$ or $B \geq A$.

SV 5. It is false that, for every x, x' : $x \succeq x'$.

SV 6. Suppose it false that $\mathbf{f} \succeq \mathbf{g}$; then, for every x , there is a (finite) partition of S such that, if \mathbf{g}' agrees with \mathbf{g} and \mathbf{f}' agrees with \mathbf{f} except on an arbitrary element of the partition, \mathbf{g}' and \mathbf{f}' being equal to $x(B)$ there, then it will be false that $\mathbf{f}' \succeq \mathbf{g}$ or $\mathbf{f} \succeq \mathbf{g}'$.

SV 7. If $\mathbf{f} \succeq \mathbf{g}(s)$ given B for every $s \in B$, then $\mathbf{f} \succeq \mathbf{g}$ given B .

From this set of axioms Savage derives the following theorem.

Theorem 5. *There is a real-valued function of consequences u , and a probability function p such that:*

$$(1) \mathbf{f} \succeq \mathbf{g} \text{ if and only if } \int_s [u(\mathbf{f}(s)) \cdot p(s)] ds > \int_s [u(\mathbf{g}(s)) \cdot p(s)] ds.$$

Furthermore, for every other function u' satisfying (1), there are numbers $c > 0$ and d such that:

$$(2) u' = c \cdot u + d.$$

In this version of Savage's theorem, property (1) is the expected utility property, commonly interpreted as a justification of the principle of maximizing expected utility. Property (2) tells us that utility is measured on an interval scale.

Axiom **SV2** is the notorious sure-thing principle, according to which "sure-thing" outcomes of a state that would be the same under any alternative act can be ignored. Several empirical studies have confirmed that people often violate this axiom, even in situations in which they are given plenty of time to think over their decision.⁴⁴ The sure-thing principle is also the primary target of Allais' paradox, which is intended to show that the sure-thing principle is equally dubious from a normative point of view.⁴⁵

Even though Savage claimed that his work should be given a normative interpretation, it is worth noting that his primary concern was to explicate his theory of subjective ("personal") probability. It was for that task he had to invent a way of measuring utility in terms of preferences between risky acts.

⁴⁴ For an overview of the empirical evidence, see Kagel and Roth 1995, Chapter 8.

⁴⁵ Allais' paradox was originally proposed in Allais 1953. For an introduction, see Resnik 1993, pp 103-5.

Other axiomatizations

The axiomatizations presented in e.g. Marschack (1950), Hershstein and Milnor (1953), Davidson et al (1957), Fishburn (1970, 1979), Harsanyi (1978, 1979), Kreps (1988, ch. 1), and Segal (1990) can be regarded as slight modifications of those proposed by von Neumann and Morgenstern (1947) and Savage (1954/72). Some of the authors mentioned here mainly had descriptive applications in mind; others have stated explicit normative intentions. A common feature of all these axiomatizations is that (i) they rely on versions of the independence axiom or sure-thing principle, and (ii) they prove a representation theorem and a corresponding uniqueness theorem for preferences among lotteries (risky acts).

Two axiomatizations with substantially different structures have been proposed by Jeffrey and Bolker (see Jeffrey 1983)⁴⁶ respectively Oddie and Milne (1991). In the axiomatization constructed by Jeffrey and Bolker, preferences, utilities and probabilities are all defined on the same set of entities, viz. a set of propositions (see Section 2.2). According to Broome (1999), two of the axioms proposed by Jeffrey and Bolker are more dubious than the rest. One is the *averaging* axiom, which “slightly resembles the independence axiom” even though “the averaging axiom is much weaker than independence”.⁴⁷ The other dubious axiom is the *impartiality* axiom, which “presupposes expected utility theory to some extent” and is thus circular.⁴⁸ Furthermore, an additional weakness of the axiomatization constructed by Jeffrey and Bolker is that its representation theorem and “uniqueness theorem” implies that there is no *unique* probability function, but rather a class of such functions. A detailed discussion of this problem, and a suggestion of how it may be handled, can be found in Joyce (1999).

The axiomatization constructed by Oddie and Milne (1991) is inter-

⁴⁶ This reference refers to Jeffrey’s work only. For Bolker’s work, which contains the mathematical details, the appropriate references can be found in Jeffrey’s book.

⁴⁷ Broome 1999, p 98.

⁴⁸ *Ibid.*

esting mainly because these authors do not aim at justifying the principle of maximizing expected utility by means of constructing a representation theorem and a corresponding uniqueness theorem. Instead, they show that if the value of an act is a function f of the values V and probabilities P of its potential outcomes, i.e. $f(V, P)$, this assumption together with three additional axioms imply that $f(V, P) = \sum_{i=1}^n v_i p_i$. A weakness is that one of the proposed axioms (Axiom V_2) resembles Savage's sure-thing principle, and can be criticized by constructing the same sort of examples.⁴⁹

5.4 Preview of the new axiomatization

As indicated in several sections of this summary, the axiomatization of the principle of maximizing expected utility constructed here differs from previous ones in certain aspects. Very briefly put, the two main innovations are that (i) this axiomatization does not rely on any version of the much criticized independence axiom or sure-thing principle, and that (ii) it is more straightforward than earlier ones, in that it does not take the form of proving a representation theorem and a corresponding uniqueness theorem (for preferences among lotteries).

As regards the first difference, both the independence axiom and the sure-thing principle can, of course, be *derived* from the principle of maximizing expected utility, and must hence be regarded as true by the adherent of this decision rule. But they are not necessary for *deriving* the principle of maximizing expected utility. This difference is subtle, but important: Instead of justifying e.g. the independence axiom by showing that a violation of it leads to sure losses in (very specific types of) sequential choices, as demonstrated by Seidenfeld⁵⁰, the acceptability of the independence axiom in the new axiomatization follows from the acceptability of the principle of maximizing expected utility, which is ultimately based on the new axioms. It can

⁴⁹ Rabinowicz 1990, p 4.

⁵⁰ Seidenfeld 1988.

be shown that the new axiomatization provides a resolution of Allais' paradox that cannot be obtained in Savage-style or von Neumann and Morgenstern-style axiomatizations.⁵¹

When it comes to the second difference, the fundamental idea in the new axiomatization is to use the concept of transformative decision rules for decomposing the principle of maximizing expected utility into a sequence of normatively reasonable subrules, instead of constructing a traditional representation theorem. In order to see how this works, note that the principle in question (**eu**) can be trivially reconstructed as a composite rule consisting of two subrules, one transformative and one effective, such that **eu** = **weigh** ◦ **max**:

weigh Transform a decision problem under risk π into a decision problem under certainty π' with the same alternative set, such that the utility of every outcome in π' equals the weighed sum of the utilities and probabilities for the corresponding outcomes in π , i.e. $\sum_{j=1}^n p(s_j) \cdot u(a_i, s_j)$.

max In a decision problem under certainty, choose an alternative act associated with the maximum utility.

The most controversial part of the **eu** rule is its transformative subrule, i.e. **weigh**, not its effective subrule **max**. However, nothing prevents the decision maker from decomposing **weigh** into new subrules, thereby splitting it into a larger number of more intuitively appealing parts. Of course, this strategy can be applied for other rules as well. That is, in order to justify the claim that a transformative decision rule **t** is normatively reasonable one can split **t** into a number of subrules **t**₁, **t**₂, . . . , **t**_n, that are justified individually.

The new axiomatization comes in two versions, presented in Paper IV and Paper V, respectively. The difference is that the first version takes a cardinal utility function for outcomes for granted, whereas in

⁵¹ See Paper IV, Section 5.

the second version such a cardinal function is derived from the axioms. That is, in the second version only qualitative data is used. The importance of this manoeuvre is discussed at length in Paper V, in which it is also compared to the traditional approach (advocated by e.g. von Neumann and Morgenstern, and Savage), which is to use the decision maker's preferences among pairs of risky acts for constructing a cardinal utility function. Below a third version of the axioms is presented, which is a slightly simplified mix of the two versions. It resembles the first version (Paper IV) in the sense that a utility function is assumed to be available to the decision maker, but the general structure of the axioms has more in common with the second version (Paper V).

A technical limitation of the new axiomatization, as presented in this thesis, is that it only applies to formal representations of decision problems with finite sets of alternative acts and finite sets of states, and presupposes that all probabilities can be expressed as rational numbers.

In the remainder of this section $\pi = \langle A, S, p, u \rangle$ is a formal decision problem under risk, in which $u(a, s)$ represents the utility associated with the outcome that follows if a is performed and s is the case. As explained in Section 3.3, $\Phi(\pi)$ is a *solution* of π if and only if (i) $\Phi(\pi) \subseteq A$ and (ii) every member in $\Phi(\pi)$ is normatively reasonable to perform. Now consider the following axioms, which all π in a set of formal decision problems Π ought to satisfy.

MP 1. *For every π , $\Phi(\pi) \neq \emptyset$.*

MP 2. *If there is an a such that $u(a, s) > u(a', s)$ for all a' and s , then $a' \notin \Phi(\pi)$.*

MP 3. *If π is transformed into π' by splitting a state into two, such that $u(a, s) = u'(a, s') = u'(a, s'')$ for all a , and $p(s) = p(s') + p(s'')$, then $\Phi(\pi) = \Phi(\pi')$.*

MP 4. *If π is transformed into π' by splitting an act into two, such that $u(a, s) = u'(a', s) = u'(a'', s)$ for all s , then, if $a \in \Phi(\pi)$ it holds*

that $\Phi(\pi') = \{a', a''\} \cup (\Phi(\pi) - a)$.

MP 5. *There is a closed interval of positive numbers $[v, w]$ such that for all $\alpha \in [v, w]$ there is a number β such that for all formal decision problems π , if $a \in \Phi(\pi)$ and $u(a, s) > u(a, s')$, it holds that $\Phi(\pi) = \Phi(\pi')$, where π' is the formal decision problem obtained from π by subtracting α from $u(a, s)$ and adding β to $u(a, s')$.*

MP 5 is the most complex and controversial axiom. It implies that in case a good outcome is diminished by a small amount, then this can be compensated for by improving a not-so-good outcome of the same act. Note that **MP 5** can be accepted even by decision makers that are aversive to utility risks, since (i) the compensation might be enormous, compared to the reduction of the other outcome, and (ii) the loss of utility is always “withdrawn” from an outcome that is better than the outcome to which the compensation is added.

Taken together **MP 1–MP 5** imply that the solution of a formal decision problem under risk is the set of acts having the highest expected utility.

Theorem 6. *Let **MP 1–MP 5** hold for a set Π of formal decision problems under risk. Then, $\Phi(\pi) = \mathbf{eu}(\pi)$ for all $\pi \in \Pi$.*

At this point it might be objected that von Neumann and Morgenstern, Savage, and their successors, are able to *define* the concept of utility in terms of preferences among risky acts (lotteries). Hence, those theories succeed in giving empirical content to the concept of utility. In fact, some people may even say that this is the main accomplishment in von Neumann and Morgenstern’s and Savage’s work. However, in the axiomatization proposed here no empirical content seems to be given to the concept of utility. So what does it *mean* in this approach that the utility of an outcome has some specified numerical value?

The reply to this objection starts with two interconnected remarks. First, let me distinguish the (partly empirical) problem of finding a

formal decision problem that corresponds in the right way to the decision problem under consideration, and that includes a reasonable utility function, from the normative problem of how to solve a formal decision problem. In von Neumann and Morgenstern's and Savage's axiomatizations the utility function is derived from preferences among risky acts. This contradicts the idea that one should separate the process in which a decision problem is modelled in a formal representation from the process of choosing which act to perform from the acts listed in that representation.

Secondly, it seems odd from a linguistic point of view to say that the *meaning* of utility has anything to do with preferences among risky alternatives. Even a decision maker who believes that he lives in a deterministic world, i.e. a world in which every act has a well-determined outcome, can meaningfully say that the utility of one certain outcome is higher (and *how much* higher it is) than the utility of another certain outcome, associated with another alternative act. In everyday contexts the concept of utility has no conceptual link to the concept of risk. Hence, it seems inadequate to apply a technical notion of utility that presupposes such a link, at least in normative discussions. (Perhaps it might be fruitful for descriptive purposes.)

However, the main reply to the objection raised above is that a precise and useful notion of utility has in fact been obtained in Paper V, by claiming that the utility of outcomes form an extensive structure. The notion of utility defended in Paper V implies, among other things, that utility has the same measurement theoretical properties as e.g. mass and length (even though there are many other important differences between these entities). Of course, some people would immediately reply that this manoeuvre makes the concept of utility too blunt. But others deny this, e.g. hedonistic utilitarians.⁵² In my view, the notion of utility advocated in this thesis is at least as reasonable as von Neumann and Morgenstern's and Savage's, especially since it

⁵² See, for example, Torbjörn Tännsjö 1998.

does not *define* utility in terms of preferences among risky acts. This point is discussed more in detail in Paper V.

5.5 A comprehensive and unified decision theory

At the outset of Chapter 1 it was indicated that the decision theory spelled out in this thesis is intended to be comprehensive and unified. Comprehensiveness was taken to mean that the theory should cover both phases of rational decision making, and unification meant that transformative decision rules should be used in both phases. Below I briefly summarize by what means these goals have been reached.

Very briefly put, it has been claimed that decision makers should start with an initial representation of a decision problem that comprises as many acts, states, probability and utility functions as possible. This initial representation should then be improved by applying a sequence of transformative decision rules that satisfy the three conditions stated in Paper I and II (see Section 4.1 and 4.2). This typically (but not always) forces the decision maker to make sure that no acts or states are left out, and that the elements in the set of acts are alternative acts, and so on. In case the set of probability functions P was empty in the initial representation it might now have been expanded by one element, e.g. by applying the principle of insufficient reason. Because of the results reported in Paper II (see Section 4.2) it does not matter in which order the transformative rules are applied; and it is always better to apply all transformative decision rules that are available and satisfy these conditions. The output of this process is a formal representation that is *non-defeated* in the sense that no alternative representation is better.

When a non-defeated formal representation has been reached an effective decision rule should be applied, e.g. the principle of maximizing expected utility. However, as shown in Paper IV and V (see Section 5.4) the principle of maximizing expected utility can be decomposed

into a sequence of several different transformative and one single effective decision rule. In this decomposed version of the principle of maximizing expected utility the transformative decision rules are the bearers of normative force, i.e. the transformative decision rules determine which acts will eventually be chosen, since the single effective rule that is applied is the rather innocent **max** rule, which can in turn be decomposed into the normatively inert **pick** rule, as shown in Section 3.2. Consequently, in the decision theory proposed here the transition from the first to the second phase of rational decision making is “invisible” from the decision maker’s point of view.

6 Comments on Papers I–VII

The seven papers that constitute the bulk of this thesis were written at different occasions and for different audiences. This explains why the papers contain some minor terminological discrepancies.

Since the first papers were written I have changed my mind concerning a few issues. In this chapter I briefly recapitulate and comment on the most central issues in each paper, and indicate what could have been done otherwise.

Paper I: Transformative Decision Rules

In Paper I the distinction between transformative and effective decision rules is introduced and explained. The focus is on some fundamental properties of transformative rules.

It is argued that no transformative decision rule is normatively reasonable unless it satisfies (or can be restated as a rule that satisfies) two properties called *convergence* respectively *iterativity*. A transformative rule is convergent just in case there is a finite number n such that the same result is obtained in case the rule is applied n and $n + 1$ times. It is iterative just in case this holds for all $n \geq 1$. It is noted that for infinite sets of decisions problems, the assumptions needed for getting the above-mentioned results, are rather strong.

In Paper I, I mention two values that together determine if the application of a transformative decision rule to a formal representation is normatively reasonable or not, viz. accuracy and simplicity. In Paper II and III a list of four values is proposed, namely realizability,

completeness, relevance, and simplicity. After Paper I was written I have come to believe that the latter list of values is more plausible than the first.

Paper II: Transformative Decision Rules, Permutability, and Non-Sequential Framing

The topic of Paper II is the permutability of transformative decision rules. A set of transformative decision rules is, roughly put, permutable just in case it does not matter in which order the rules are applied. Two notions of permutability, viz. weak and strong permutability, are distinguished. It is argued that in order to be normatively reasonable, sets of transformative decision rules have to satisfy a number of structural conditions that together imply weak permutability. This formal result gives support to a non-sequential theory of framing, i.e. a theory which prescribes no uniform order in which different steps in the framing process have to be performed.

In Paper I it was assumed that a transformative decision rule is normatively reasonable just in case it strictly improves the formal decision problem, given that it is changed by the rule in question. In Paper II a slightly weaker criterion is proposed, which is discussed at length in that paper, as well in Section 4.1 of this summary.

The term ‘framing’ is used throughout Paper II. It can be taken to be synonymous with the somewhat clumsy term ‘the process in which a decision problem is modelled in a formal representation’.

Paper III: Rival Representations of Decision Problems

The point of departure in Paper III is the insight that two or more formal representations can correspond optimally to a given decision problem, in the sense that no alternative representation is better. Such formal decision problems are called rival representations.

Sometimes an act will be ranked as best by an effective decision rule, e.g. the principle of maximizing expected utility, when evalu-

ated in one optimal representation of a given decision problem, but ranked as non-best by the same effective decision rule in another optimal representation of the same decision problem. In such cases one may legitimately ask whether the act in question should be performed or not. This problem is addressed by identifying two positions, ecumenism and absolutism, and argue that the former is more plausible than the latter.

Somewhat roughly put, ecumenism is the claim that all rival representations are equally reasonable and it is indifferent which one the decision maker decides to apply. It should be observed that the ecumenic position does not imply that decision theory cannot provide decision makers with action guidance. Because if an act is judged as both optimal and non-optimal (in different rival representations), ecumenists can claim that it is *permitted* both to perform and not to perform the act in question. But in case some acts are non-optimal in every rival representation they are, according to ecumenists, *forbidden*. Analogously, an act is *obligatory* just in case it is permitted in every rival representation and no alternative act is permitted in any rival representation.

Paper IV: An Argument for the Principle of Maximizing Expected Utility

A key observation in Paper IV is that the principle of maximizing expected utility can be decomposed into a sequence of transformative and effective subrules, that can be justified individually. The six axioms suggested in Paper IV together imply that it is normatively reasonable to apply the principle of maximizing expected utility to every decision problem under risk. The proof consists in a construction of a sequence of transformative and effective decision rules, which is justified through applying an idea introduced in Paper I.

None of the axioms proposed in Paper IV imply, or can be derived from, the independence axiom or the sure-thing principle. It is argued

that this opens up the door for a resolution of the notorious Allais' paradox, which cannot be obtained with arguments relying on those axioms.

In Paper I it was assumed that it is normatively reasonable to transform a decision problem into another just in case the new decision problem is strictly better (more reasonable) than the old one. If two decision problems are of equal value, then a transformation is not normatively reasonable, according to Paper I. At first glance, it might be thought that this assumption contradicts the axioms of Paper IV, saying that transformations are normatively reasonable even in case they do not yield strictly better decision problems, only decision problems of equal value. I now believe that the assumption in Paper I only holds for entire principles that are actually applied by decision makers, e.g. the principle of insufficient reason, not for axioms used for decomposing such rules. The assumption of Paper I is justified because real-life decision makers may make mistakes if they perform unnecessary transformations. The axioms of Paper IV should mainly be used for constructing a complex decision rule from several subrules. The resulting final rule can thereafter be used in real-life applications, and its transformative part (**weigh**) obviously satisfies the assumption stated in Paper I.

In Paper II a number of conditions are stated that I claim hold for all transformative decision rules that are normatively reasonable. Do the transformative rules constructed in Paper IV satisfy these conditions? My current standpoint is that that is the case: All transformations carried out by the three transformative decision rules derived from the axioms return formal representations that are equally good representations of the decision problem in question, and that are monotone and iterative when applied to all formal decision problems under risk. However, whether this claim of mine is correct or not is a question of values, so I would not be surprised if some people disagree.

Paper V: From Outcomes to Acts: A Non-Standard Axiomatization of the Expected Utility Principle

In Paper V an improved version of the axiomatization proposed in Paper IV is introduced. The most significant improvement is that no numerical utility function is taken for granted. Instead, the utility function is derived from purely qualitative information.

In Paper V it is argued that attempts to justify the principle of maximizing expected utility by means of proving a representation theorem and a uniqueness theorem for preferences among *lotteries* are of little normative relevance, since such axiomatizations offer no (or almost no) action-guidance. To briefly summarize the argument, remember that in such an *ex post* approach, preferences among risky alternatives are used as input data in the construction of a utility function (and in some cases also a probability function). However, if the decision maker already knows what risky alternatives he or she prefers, what is then the point in constructing a utility function? That procedure does not seem to yield any additional action-guidance.

In the subsequent sections of Paper V an *ex ante* argument for the principle of maximizing expected utility is constructed. It is shown that preferences among (risk-free) outcomes, in conjunction with a number of rather weak axioms, generates preferences for risky acts.

An important difference between the axiomatization proposed in Paper V and that proposed in Paper IV is that the latter deals with *decision rules*, whereas the former is concerned with *formal decision problems*. That is, in Paper IV the main theorem states that a certain decision rule (the principle of maximizing expected utility) is normatively reasonable, whereas in Paper V the main theorem states that the solution of every formal decision problem under risk is identical to those acts recommended by the principle of maximizing expected utility.

According to the rule-based axiomatization the ultimate aim of normative decision theory is to justify a particular decision rule, and ac-

According to the act-based axiomatization the ultimate aim of normative decision theory is to justify a particular set of acts. The difference is subtle, but seems to be of some general philosophical interest: do we wish to justify a certain way of making decisions, or just certain decisions? Or slightly differently put: Do we have to do the right thing for the right reasons, or is it sufficient that we just do the right thing, no matter what the reasons are? It is fruitful to consider this question in light of the standard account of knowledge (knowledge as true, justified, belief): Do we have to believe what is true for the right reasons, or is it enough that we just make sure that we believe what is true, no matter what the reasons are? Since I am not sure about the answer to the first of these problems I have, as explained above, developed two alternative axiomatizations that take care of both possibilities.

Paper VI: The Limits of Catastrophe Aversion

In the sixth paper, Paper VI, the management of catastrophe-risks is discussed from a theoretical point of view. First, the concept of a catastrophe is informally and formally defined. Thereafter, a number of seemingly weak desiderata for catastrophe averse decision rules are introduced. However, it is shown that the proposed desiderata are in fact mutually inconsistent. In consequence of this result, it is argued that the *rigid* form of catastrophe aversion articulated by e.g. the maximin rule, the maximum probable loss rule, (some versions of) the precautionary principle, and the rule proposed in Ekenberg et al (2001) should be given up. An alternative form of *non-rigid* catastrophe aversion is considered.

Paper VII: The Ethics of Extreme Risks

In the last paper, Paper VII, I discuss social decision-making involving extreme risks. By an extreme risk I mean a potential outcome of an act for which the probability is low, but whose (negative) value is high. Extreme risks are often discussed when new technologies are

introduced into society. Nuclear power and genetic engineering are two well-known examples.

At the end of Paper VII, I advocate a composite decision rule, (**dm** ◦ **pp** ◦ **eu**) that is catastrophe averse in the sense explained in Paper VI. In order to make the rule advocated in Paper VII consistent with the conclusion of Paper VI, the version of the precautionary principle discussed in Paper VII has to be improved along the lines outlined at the end of paper VI, i.e. it has to be stated in a non-rigid form.

Furthermore, in order to see how the proposed rule can be consistent with the axiomatic defense of the expected utility principle presented in Papers IV and V, note that **dm** and **pp** are used in the framing of the decision problem. That is, these rules are used for obtaining an optimal representation of a decision problem under risk, to which the **eu** is thereafter applied.

The term 'ecumenism' is used in the last section of Paper VII. Note that this term does not have the same meaning in this paper as in Paper III.

Appendix

Proof of Theorem 1

Let π be an arbitrary element in Π .

- | | | |
|------|---|--|
| (1) | $\mathbf{u}(\pi) \succeq (\pi)$ | Cond. 1 |
| (2) | $(\mathbf{u} \circ \mathbf{t})(\pi) \succeq \mathbf{t}(\pi)$ | (1), Cond. 2 |
| (3) | $(\mathbf{u} \circ \mathbf{t} \circ \mathbf{u})(\pi) \succeq (\mathbf{t} \circ \mathbf{u})(\pi)$ | (2), Cond. 2 |
| (4) | $(\mathbf{t} \circ \mathbf{u})(\pi) \sim (\mathbf{t} \circ \mathbf{u} \circ \mathbf{u})(\pi)$ | Cond. 3, note that $\mathbf{t}(\pi) \in \Pi$ |
| (5) | $(\mathbf{u} \circ \mathbf{t} \circ \mathbf{u})(\pi) \succeq (\mathbf{t} \circ \mathbf{u} \circ \mathbf{u})(\pi)$ | (3), (4) |
| (6) | $(\mathbf{u} \circ \mathbf{t} \circ \mathbf{u} \circ \mathbf{t})(\pi) \succeq (\mathbf{t} \circ \mathbf{u} \circ \mathbf{u} \circ \mathbf{t})(\pi)$ | (5), Cond. 2 |
| (7) | $(\mathbf{u} \circ \mathbf{t})(\pi) \succeq (\mathbf{t} \circ \mathbf{u})(\pi)$ | (6), Cond. 2 applied twice |
| (8) | $\mathbf{t}(\pi) \succeq (\pi)$ | Cond. 1 |
| (9) | $(\mathbf{t} \circ \mathbf{u})(\pi) \succeq \mathbf{u}(\pi)$ | (8), Cond. 2 |
| (10) | $(\mathbf{t} \circ \mathbf{u} \circ \mathbf{t})(\pi) \succeq (\mathbf{u} \circ \mathbf{t})(\pi)$ | (9), Cond. 2 |
| (11) | $(\mathbf{u} \circ \mathbf{t})(\pi) \sim (\mathbf{u} \circ \mathbf{t} \circ \mathbf{t})(\pi)$ | Cond. 3, note that $\mathbf{u}(\pi) \in \Pi$ |
| (12) | $(\mathbf{t} \circ \mathbf{u} \circ \mathbf{t})(\pi) \succeq (\mathbf{u} \circ \mathbf{t} \circ \mathbf{t})(\pi)$ | (10), (11) |
| (13) | $(\mathbf{t} \circ \mathbf{u} \circ \mathbf{t} \circ \mathbf{u})(\pi) \succeq (\mathbf{u} \circ \mathbf{t} \circ \mathbf{t} \circ \mathbf{u})(\pi)$ | (12), Cond. 2 |
| (14) | $(\mathbf{t} \circ \mathbf{u})(\pi) \succeq (\mathbf{u} \circ \mathbf{t})(\pi)$ | (13), Cond. 2 applied twice |
| (15) | $(\mathbf{t} \circ \mathbf{u})(\pi) \sim (\mathbf{u} \circ \mathbf{t})(\pi)$ | (7), (14) |

Proof of Theorem 2

Let π be an arbitrary element in Π .

- | | | |
|-----|---|--|
| (1) | $(\mathbf{u} \circ \mathbf{t})(\pi) \succeq (\mathbf{t} \circ \mathbf{u})(\pi)$ | Theorem 1 |
| (2) | $(\mathbf{u} \circ \mathbf{t} \circ \mathbf{t})(\pi) \succeq (\mathbf{t} \circ \mathbf{u} \circ \mathbf{t})(\pi)$ | (1), Cond. 2 |
| (3) | $(\mathbf{u} \circ \mathbf{t})(\pi) \sim (\mathbf{u} \circ \mathbf{t} \circ \mathbf{t})(\pi)$ | Cond. 3, note that $\mathbf{u}(\pi) \in \Pi$ |
| (4) | $(\mathbf{u} \circ \mathbf{t})(\pi) \succeq (\mathbf{t} \circ \mathbf{u} \circ \mathbf{t})(\pi)$ | (2), (3) |
| (5) | $(\mathbf{t} \circ \mathbf{u})(\pi) \succeq (\mathbf{u} \circ \mathbf{t})(\pi)$ | Theorem 1 |
| (6) | $(\mathbf{t} \circ \mathbf{u} \circ \mathbf{t})(\pi) \succeq (\mathbf{u} \circ \mathbf{t})(\pi)$ | (5), Cond. 1 |
| (7) | $(\mathbf{t} \circ \mathbf{u} \circ \mathbf{t})(\pi) \sim (\mathbf{u} \circ \mathbf{t})(\pi)$ | (4), (6) |

Proof of Observation 1

The proofs of Parts 1 and 2 are omitted. For part 3, let π be an arbitrary element in Π .

- (1) $(\mathbf{t} \circ \mathbf{u} \circ \mathbf{t} \circ \mathbf{u})(\pi) \sim (\mathbf{t} \circ \mathbf{u} \circ \mathbf{t} \circ \mathbf{u})(\pi)$ reflexivity (Sect. 4.1)
- (2) $(\mathbf{t} \circ \mathbf{u} \circ \mathbf{u} \circ \mathbf{t})(\pi) \sim (\mathbf{t} \circ \mathbf{u} \circ \mathbf{t} \circ \mathbf{u})(\pi)$ (1), Theorem 1
- (3) $(\mathbf{t} \circ \mathbf{u} \circ \mathbf{t})(\pi) \sim (\mathbf{t} \circ \mathbf{u} \circ \mathbf{t} \circ \mathbf{u})(\pi)$ (2), Cond. 3
- (4) $(\mathbf{t} \circ \mathbf{t} \circ \mathbf{u})(\pi) \sim (\mathbf{t} \circ \mathbf{u} \circ \mathbf{t} \circ \mathbf{u})(\pi)$ (3), Theorem 1
- (5) $(\mathbf{t} \circ \mathbf{u})(\pi) \sim (\mathbf{t} \circ \mathbf{u} \circ \mathbf{t} \circ \mathbf{u})(\pi)$ (4), Cond. 3

Proof of Observation 2

We prove this by induction: It follows from Theorem 1 that the claim holds in case \mathbf{T}' has 2 elements. (In case \mathbf{T}' has only one element the claim is trivially true, because of reflexivity.) In order to prove the inductive step, suppose that the claim holds in case \mathbf{T}' has n ($n \geq 2$) elements. Let \mathbf{t} be element $n + 1$, and let \mathbf{F} be a sequence of v elements and let \mathbf{G} be a sequence of w elements; $v + w = n$.

We need to show that $(\mathbf{t} \circ \mathbf{F} \circ \mathbf{G})(\pi) \sim (\mathbf{F} \circ \mathbf{t} \circ \mathbf{G})(\pi) \sim (\mathbf{F} \circ \mathbf{G} \circ \mathbf{t})(\pi) \sim (\mathbf{t} \circ \mathbf{G} \circ \mathbf{F})(\pi) \sim (\mathbf{G} \circ \mathbf{t} \circ \mathbf{F})(\pi) \sim (\mathbf{G} \circ \mathbf{F} \circ \mathbf{t})(\pi)$. First consider the case in which \mathbf{F} and \mathbf{G} have a non-zero number of elements, i.e. $v, w \neq 0$. Note that the number of elements in $(\mathbf{t} \circ \mathbf{G})$ is $\leq n$. Hence, since the theorem was assumed to hold for up to n elements and $\mathbf{F}(\pi) \in \Pi$, it follows that $(\mathbf{F} \circ \mathbf{t} \circ \mathbf{G})(\pi) \sim (\mathbf{F} \circ \mathbf{G} \circ \mathbf{t})(\pi)$. We also use the fact that the number of elements in $(\mathbf{t} \circ \mathbf{F})$ is $\leq n$: By applying Observation 1, Part 2, repeatedly we find that if $\pi \sim \pi'$, then $\mathbf{G}(\pi) \sim \mathbf{G}(\pi')$ for all $\pi, \pi' \in \Pi$. Therefore, since the theorem was assumed to hold for n elements, it follows that $(\mathbf{t} \circ \mathbf{F} \circ \mathbf{G})(\pi) \sim (\mathbf{F} \circ \mathbf{t} \circ \mathbf{G})(\pi)$. So far we have shown (since \sim is transitive) that $(\mathbf{t} \circ \mathbf{F} \circ \mathbf{G})(\pi) \sim (\mathbf{F} \circ \mathbf{t} \circ \mathbf{G})(\pi) \sim (\mathbf{F} \circ \mathbf{G} \circ \mathbf{t})(\pi)$; by applying an analogous argument we find that $(\mathbf{t} \circ \mathbf{G} \circ \mathbf{F})(\pi) \sim (\mathbf{G} \circ \mathbf{t} \circ \mathbf{F})(\pi) \sim (\mathbf{G} \circ \mathbf{F} \circ \mathbf{t})(\pi)$. Finally, since the number of elements in $(\mathbf{F} \circ \mathbf{G})$ is $= n$, it follows, together with the assumption that the theorem holds for n elements, that $(\mathbf{t} \circ \mathbf{F} \circ \mathbf{G})(\pi) \sim (\mathbf{t} \circ \mathbf{G} \circ \mathbf{F})(\pi)$. Since \sim is transitive it follows that all six permutations of \mathbf{t}, \mathbf{F} and \mathbf{G} are of equal value.

The second case, in which the number of elements in either \mathbf{F} or \mathbf{G} is zero, is trivial, since we have shown above that $(\mathbf{t} \circ \mathbf{F} \circ \mathbf{G})(\pi) \sim (\mathbf{F} \circ \mathbf{G} \circ \mathbf{t})(\pi)$.

Proof of Theorem 3

Let $\mathbf{C} = \mathbf{B} - \mathbf{A}$. From Observation 2 it follows that for every $p_b(\pi)$ there is a permutation $p_c(\pi)$ such that $p_a \circ p_c(\pi) \sim p_b(\pi)$. Hence, because of Condition 1, $p_b(\pi) \succeq p_a(\pi)$.

Proof of Theorem 4

See von Neumann and Morgenstern 1947, pp 618-632, for the original proof. A more accessible proof can be found in Kreps 1988.

Proof of Theorem 5

See Savage 1954/72, Chapter 3.

Proof of Theorem 6

Theorem 6 is easier to prove if one first prove the following lemma.

Lemma 1. (*Equal Trade-Off*) *There is a closed interval of positive numbers $[v, w]$ such that for all $\alpha \in [v, w]$ and for all formal decision problems π , if $a \in \Phi(\pi)$, s and s' are equi-probable, and $u(a, s) > u(a, s')$, it holds that $\Phi(\pi) = \Phi(\pi')$, where π' is the formal decision problem obtained from π by subtracting α from $u(a, s)$ and adding α to $u(a, s')$.*

We are going to prove Lemma 1 by showing that $\alpha = \beta$ whenever MP 5 is applied to a formal decision problem with equi-probable states. Assume for *reductio* that $\alpha \neq \beta$, and let a be an act and let s and s' be two equi-probable states in π . Let us first consider the case in which s and s' are the only states in π , and $a \in \Phi(\pi)$. Let $u(a, s) = u_1$ and let $u(a, s') = u_2$; $u_1 > u_2$. By applying MP 4 twice, we find that π can be transformed into a formal decision problem π' with three new (identical) acts a', a'', a''' such that $\Phi(\pi') = \{\Phi(\pi) - a\} \cup \{a' \cup a'' \cup a'''\}$. (See figure below.)

We now apply MP 5 to a'' and a''' in π' exactly k times with some α' and β' , such that $0 < u_1 - k\alpha' - [u_2 + k\beta'] < \alpha$. (It might hold that there are no k, α', β' that satisfy this inequality, e.g. if $\beta' \gg \alpha'$. In that case, let $k = 1$, $\alpha' = \alpha$, $\beta' = \beta$, and ignore the next sentence.) Then we apply MP 5 to a''' one more time, this time subtracting α respectively adding β . It now holds that $u(a''', s) < u(a''', s')$ in π'' . Finally, we apply MP 5 again, but this time we subtract α from $u(a''', s')$ and add β to $u(a''', s)$.

π'			π''		
	s	s'		s	s'
...			...		
a'	u_1	u_2	a'	u_1	u_2
a''	u_1	u_2	a''	$u_1 - k\alpha'$	$u_2 + k\beta'$
a'''	u_1	u_2	a'''	$u_1 - k\alpha' - \alpha + \beta$	$u_2 + k\beta' + \beta - \alpha$
...			...		

We initially assumed that $\alpha \neq \beta$. Hence, $-\alpha + \beta \neq 0$. Thus, by MP 2, $a'' \notin \Phi(\pi'')$ or $a''' \notin \Phi(\pi'')$, which contradicts the result that $\{a', a'', a'''\} = \Phi(\pi') = \Phi(\pi'')$. Consequently, the initial assumption must be false.

Let us now consider the case in which $a \notin \Phi(\pi)$. Then, by MP 1, there is some other act b such that $b \in \Phi(\pi)$. Furthermore, since β is a function of α , it follows that α and β can be applied to $u(b, s)$ and $u(b, s')$ in the same way as above, yielding the analogous contradiction.

Finally, in order to handle the general case with more than two equi-probable states, it is sufficient to note that for every state $v \in \{S - \{s \cup s'\}\}$, it holds that $u(a, v) = u(a', v) = u(a'', v) = u(a''', v)$. Hence, the manoeuvre from π to π''' can be carried out in a similar way as above.

Q.E.D. for Lemma 1.

We are now in a position to prove Theorem 6. The proof proceeds by the application of a series of transformative decision rules to an arbitrary element π in Π . We wish to show that $\Phi(\pi) = \mathbf{eu}(\pi)$, where $\mathbf{eu}(\pi) = \{a : \sum_{s \in S} u(a, s) \cdot p(s) \geq \sum_{s \in S} u(a', s) \cdot p(s)$ for all $a' \in A\}$.

Step 1: First, we apply to π the rule \mathbf{ss}^* (split of states), which we define as the following rule \mathbf{ss} iterated n times for some n such that $\mathbf{ss}^n(\pi) = \mathbf{ss}^{n+1}(\pi)$, where \mathbf{ss}^n denotes the iteration of \mathbf{ss} n times. Let d equal the inverse of the denominator of the rational number $p(s_1) \cdot p(s_2) \dots \cdot p(s_n)$.

$$\mathbf{ss}(\pi) \begin{cases} \pi & \text{if } p(s_1) = \dots = p(s_n) \\ \langle A', S', p', u' \rangle & \text{otherwise, with for all } p, q, r : \end{cases}$$

1. $A' = A$
2. $S' = (S - s) \cup s^\alpha \cup s^\beta$, where s is some state in π for which there is some other state $s_a \in S$ in π with $p(s) > p(s_a)$
3. $p'(s_p) = p(s_p)$ whenever $s_p \in S - s$
 $p'(s^\alpha) = p(s) - d$
 $p'(s^\beta) = d$
4. $u'(a_q, s_r) = u(a_q, s_r)$ whenever $s_r \in S - s$
 $u'(a_q, s_r) = u(a_q, s)$ otherwise

Rule \mathbf{ss}^* will converge in such way that all states receive probability d in $\mathbf{ss}^*(\pi)$, because every time we apply \mathbf{ss} exactly one state is split into two, of which at least one has a probability equal to d and the other one has a probability $\geq nd$ for some positive integer n . Since rule \mathbf{ss}^* is obtained from an iterated application of MP 3, it follows that $\Phi(\pi) = \Phi(\mathbf{ss}^*(\pi))$. Furthermore, $\mathbf{eu}(\pi) = \mathbf{ss}^* \circ \mathbf{eu}(\pi)$ because $p(s) \cdot u = (p(s) - d) \cdot u + d \cdot u$ for all $p(s), d, u$.

Step 2: We now apply to $\mathbf{ss}^*(\pi)$ the rule \mathbf{to}^* (trade-off), which we define as the following rule \mathbf{to} iterated n times for some n such that $\mathbf{to}^n(\pi) = \mathbf{to}^{n+1}(\pi)$. Let $high(\pi, a)$ be a function that returns an ordered pair $\langle a, s_j \rangle$ such that $u(a, s_j) \geq u(a, s_k)$ for every k , and $low(\pi, a)$ a function that returns an ordered pair $\langle a, s_j \rangle$ such that $u(a, s_j) \leq u(a, s_k)$ for every k . Let ε be an arbitrary small positive non-zero number such that (i) $0 < \varepsilon \leq \alpha$, and (ii) for every $u(a_i, s_j)$ there is an even number t such that $\varepsilon \cdot t = u(a_i, s_j)$.

$$\mathbf{to}(\pi) \begin{cases} \pi & \text{if } u(a_o, s_p) = u(a_o, s_q) \text{ for all } o, p, q \\ \langle A', S', p', u' \rangle & \text{otherwise, with :} \end{cases}$$

1. $A' = A$
2. $S' = S$
3. $p' = p$
4. For some a such that $h = high(\pi, a), l = low(\pi, a)$ and $u(h) \neq u(l)$, for all r, t, v, w :
 - $u'(a, s_r) = u(a, s_r) - \varepsilon$ if and only if $\langle a, s_r \rangle = h$
 - $u'(a, s_t) = u(a, s_t) + \varepsilon$ if and only if $\langle a, s_t \rangle = l$
 - $u'(a_v, s_w) = u(a_v, s_w)$ for all $\langle a_v, s_w \rangle \neq h, l$

Since rule \mathbf{to}^* is obtained from an iterated application of Lemma 1 (equal trade-off) and $p(s_1) = \dots = p(s_n)$ in $\mathbf{ss}^*(\pi)$, it holds that $\Phi(\mathbf{ss}^*(\pi)) = \Phi(\mathbf{ss}^* \circ \mathbf{to}^*(\pi))$. It follows from the definition of ε above that \mathbf{to}^* will converge such that $u(a_j, s_k) = u(a_j, s_l)$ for all j, k, l in the formal decision problem $\mathbf{ss}^* \circ \mathbf{to}^*(\pi)$. Furthermore, $\mathbf{eu}(\pi) = \mathbf{ss}^* \circ \mathbf{to}^* \circ \mathbf{eu}(\pi)$ because $p \cdot u_1 + p \cdot u_2 = p \cdot (u_1 - \varepsilon) + p \cdot (u_2 + \varepsilon)$ for all p, u_1, u_2 and ε .

Step 3: We now apply to $\mathbf{ss}^* \circ \mathbf{to}^*(\pi)$ the rule \mathbf{ms}^* (merger of states), which we define as the following rule \mathbf{ms} iterated n times for some n such that $\mathbf{ms}^n(\pi) = \mathbf{ms}^{n+1}(\pi)$. Let s_a and s_b be two arbitrary states in S .

$$\mathbf{ms}(\pi) \begin{cases} \pi & \text{if } u(a_o, s_p) = u(a_o, s_q) \text{ for all } o, p, q \\ \langle A', S', p', u' \rangle & \text{otherwise, with :} \end{cases}$$

1. $A' = A$
2. $S' = (S - s_a - s_b) \cup s_z$, where s_z is a new state
3. $p'(s_p) = p(s_p)$ whenever $s_p \in (S - s_a - s_b)$
 $p'(s_z) = p(s_a) + p(s_b)$
4. $u'(a_q, s_r) = u(a_q, s_r)$ whenever $s_r \in (S - s_a - s_b)$
 $u'(a_q, s_z) = u(a_q, s_a)$

Since rule \mathbf{ms}^* is obtained from an iterated application of MP 3, used “backwards”, and $u(a_p, s_q) = u(a_p, s_r)$ for all p, q, r in $\mathbf{ss}^* \circ \mathbf{to}^*(\pi)$, it holds that $\Phi(\mathbf{ss}^* \circ \mathbf{to}^*(\pi)) = \Phi(\mathbf{ss}^* \circ \mathbf{to}^* \circ \mathbf{ms}^*(\pi))$. Note that rule \mathbf{ms}^* will converge such that $S = \{s\}$ in the formal decision problem $\mathbf{ss}^* \circ \mathbf{to}^* \circ \mathbf{ms}^*(\pi)$, i.e. that $\mathbf{ss}^* \circ \mathbf{to}^* \circ \mathbf{ms}^*(\pi)$ is a decision problem under certainty. Furthermore, $\mathbf{ss}^* \circ \mathbf{to}^* \circ \mathbf{ms}^* \circ \mathbf{eu}(\pi) = \mathbf{ss}^* \circ \mathbf{to}^* \circ \mathbf{eu}(\pi)$ for obvious arithmetical reasons.

Step 4: We now apply to $\mathbf{ss}^* \circ \mathbf{to}^* \circ \mathbf{ms}^*(\pi)$ the rule \mathbf{max} , which we define as follows:

$$\mathbf{max}(\pi) = \{a_p : u(a_p, s) \geq u(a_q, s) \text{ for all } q\}$$

Since $\mathbf{ss}^* \circ \mathbf{to}^* \circ \mathbf{ms}^*(\pi)$ is a decision problem under certainty, as noted in Step 3, it follows from MP 2 (dominance), together with MP 1 and MP 3, that:

$$\Phi(\mathbf{ss}^* \circ \mathbf{to}^* \circ \mathbf{ms}^*(\pi)) = \mathbf{ss}^* \circ \mathbf{to}^* \circ \mathbf{ms}^* \circ \mathbf{max}(\pi) \quad (1)$$

In every decision problem under certainty π_c it holds that $\mathbf{max}(\pi_c) = \mathbf{eu}(\pi_c)$. Hence:

$$\mathbf{ss}^* \circ \mathbf{to}^* \circ \mathbf{ms}^* \circ \mathbf{max}(\pi) = \mathbf{ss}^* \circ \mathbf{to}^* \circ \mathbf{ms}^* \circ \mathbf{eu}(\pi) \quad (2)$$

Furthermore, from Steps 1-3 it follows that:

$$\mathbf{ss}^* \circ \mathbf{to}^* \circ \mathbf{ms}^* \circ \mathbf{eu}(\pi) = \mathbf{eu}(\pi) \quad (3)$$

$$\Phi(\mathbf{ss}^* \circ \mathbf{to}^* \circ \mathbf{ms}^*(\pi)) = \Phi(\pi) \quad (4)$$

From (1)-(4) we can conclude that $\Phi(\pi) = \mathbf{eu}(\pi)$ for any arbitrary decision problem π in Π .

Grand Q.E.D.

Bibliography

- 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.
- Arnauld, A. and P. Nicole (1662/1996) *Logic or the Art of Thinking*, 5th ed., Translated and edited by Jill Vance Buroker, Cambridge University Press.
- Arrow, K. J. (1951) 'Alternative Approaches to the Theory of Choice in Risk-Taking Situations', *Econometrica* **19**, 404–437.
- Bergström, L. (1966) *The Alternatives and Consequences of Actions*, Almqvist & Wiksell International.
- Bernoulli, D. (1738/1954) 'Specimen Theoriae Novae de Mensura Sortis', *Commentari Academiae Scientiarum Imperialis Petropolitanae* **5**, 175–192. Translated as: 'Expositions of a New Theory on the Measurement of Risk', *Econometrica* **22**, 23–36.
- Brim, O. G. (1962) *Personality and Decision Processes*, *Studies in the Social Psychology of Thinking*, Stanford.
- Broome, J. (1999) *Ethics out of Economics*, Cambridge University Press.
- Carlsson, E. (1995) *Consequentialism Reconsidered*, Kluwer Academic Publishers.
- Clemen, R. T. (1991) *Making hard decisions: an introduction to decision analysis*, PWS-Kent Publishing Company.
- Davidson, D., Suppes, P., and Siegel, S. (1957) *Decision Making: An Experimental Approach*, Stanford University Press.

- Ekenberg L., Boman M., Linnerooth-Bayer J. (2001) 'General Risk Constraints', *Journal of Risk Research* **4**, 31–47.
- Fishburn, P. (1970) *Utility Theory for Decision Making*, John Wiley and Sons. Reprinted by Krieger Press 1979.
- Gärdenfors P. and Nils-Eric Sahlin (1982), Unreliable probabilities, risk taking, and decision making, pp 323–334 in Gärdenfors and Sahlin (1988).
- Gärdenfors P. and N-E. Sahlin, eds. (1988), *Decision, probability, and utility*, Cambridge University Press.
- Hansson, S. O. (2001) *The Structure of Values and Norms*, Cambridge University Press.
- Harsanyi, J. C. (1978) 'Bayesian Decision Theory and Utilitarian Ethics', *American Economic Review, Papers & Proc.* **68**, 223–228.
- Harsanyi, J. C. (1979) 'Bayesian Decision Theory, Rule Utilitarianism, and Arrow's Impossibility Theorem', *Theory and Decision* **11**, 289–317.
- Hempel, C. G. (2000) *Selected philosophical essays*, Cambridge University Press.
- Herstein, I. N. and J. Milnor (1953) 'An axiomatic approach to measurable utility', *Econometrica* **21**, 291–297.
- Jeffrey, R. (1983) *The Logic of Decisions*, 2nd edition (significant improvements from 1st), University of Chicago Press.
- Joyce, J. M. (1999) *The foundations of causal decision theory*, Cambridge University Press.
- Kagel, J. H. and A. Roth (1995) *The Handbook of Experimental Economics*, Princeton University Press.
- Keynes, J. M. (1921) *A Treatise on Probability*, MacMillan & Co.
- Keynes, J. M. (1923) *A Tract on Monetary Reform*, MacMillan & Co.

- Kihlbom, U. (2002) *Ethical Particularism: An Essay on Moral Reasons*, Almqvist & Wiksell International.
- Kreps, D. M. (1988) *Notes on the Theory of Choice*, Westview Press.
- Levi, I. (1980) *The Enterprise of Knowledge*, MIT Press.
- Luce, D. and H. Raiffa (1957) *Games and Decisions: Introduction and Critical Survey*, Wiley and Sons.
- Milnor J.W. (1954) 'Games against nature', in Thrall et al (eds.): *Decision Processes*, Wiley & Sons, 49–60.
- von Neumann, J. and O. Morgenstern (1947) *Theory of Games and Economic Behavior*, 2nd edition, Princeton University Press. [1st ed. without utility theory.]
- Nozick, R. (1969) 'Newcomb's Problem and Two Principles of Choice', in Rescher et al (eds.), *Essays in Honor of Carl G. Hempel*, Reidel, Dordrecht, 114–146.
- Oddie, G. and P. Milne (1991) 'Act and Value: Expectation and the Representability of Moral Theories', *Theoria* **57**, 42–76.
- Pollock, J. L. (2002) 'Causal Probability', *Synthese* **132**, 143–85.
- Quine, W. V. O. (1951) 'Two Dogmas of Empiricism', *The Philosophical Review* **60**, 20–43.
- Quine, W. V. O. (1992) *Pursuit of truth*, Harvard University Press.
- Rabinowicz, W. (1990) 'Rationella beslut under oskerhet eller Hur skall man handla blindo?' (in Swedish), *Filosofisk Tidskrift* **11**, 1–12.
- Rabinowicz, W. (1997) 'On Seinfeld's Criticism of Sophisticated Violations of the Independence Axiom', *Theory and Decision* **43**, 279–292.
- Ramsey, F. P. (1926) 'Truth and Probability', in *Foundations: Essays in Philosophy, Logics, and Mathematics and Economics*, Routledge & Kegan Paul 1978.
- Resnik, M. (1993) *Choices. An introduction to decision theory*, University of Minnesota Press.

- Sahlin, N-E. (1990) *The philosophy of F. P. Ramsey*, Cambridge University Press.
- Sandin, P. (1999) ‘Dimensions of the Precautionary Principle’, *Human and Ecological Risk Assessment* **5**, 889–907.
- Savage, L. J. (1954) *The Foundations of Statistics*, Wiley and Sons. 2nd edition 1972, Dover.
- Schmidt, U. (1998) *Axiomatic Utility Theory under Risk: Non-Archimedean Representations and Application to Insurance Economics*, Lecture Notes in Economics and Mathematical Systems, Springer Verlag.
- Segal, U. (1990) ‘Two-Stage Lotteries without the Reduction Axiom’, *Econometrica* **58**, 349–379.
- Seidenfeld, T. (1988) ‘Decision Theory without “Independence” or without “ordering”’: What Is the Difference?’, *Economics and Philosophy* **4**, 267–290.
- Tännsjö, T. (1998) *Hedonistic Utilitarianism*, Edinburgh University Press.
- Whipple, Chris (ed.) (1987), *De Minimis Risk*, Plenum Press.