

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

**The treatment of uncertainty in
multicriteria decision making**

University of Cape Town

The treatment of uncertainty in multicriteria decision making

Ian Durbach

University of Cape Town

This dissertation is prepared in fulfilment of the requirements for the degree of
Master of Business Science in Operational Research at the University of Cape Town.

Supervised by Professor Theo Stewart

June 2003

Acknowledgements

I am indebted to the insightful mind of Professor Theo Stewart.

University of Cape Town

Synopsis

The nature of human decision making dictates that a decision must often be considered under conditions of uncertainty. Decisions may be influenced by uncertain future events, doubts regarding the precision of inputs, doubts as to what the decision maker considers important, and many other forms of uncertainty. The multicriteria decision models that are designed to facilitate and aid decision making must therefore consider these uncertainties if they are to be effective. In this thesis, we consider the treatment of uncertainty in multicriteria decision making (MCDM), with a specific view to investigating

- the types of uncertainty that are most relevant to MCDM,
- how the uncertainties identified as relevant may be treated by various different MCDM methodologies.

We address the first of these objectives as part of the literature survey conducted in chapter 2. We identify three key categories of uncertainty.

1. Risk, relating to uncertainty about uncertain future states of the world.
2. Imprecision, relating to uncertainty about the accuracy of inputs to the decision problem.
3. Vagueness, relating to uncertainty about the scope and design of the decision problem.

In the latter stages of chapter 2 we briefly consider the use of fuzzy sets as a model of imprecision, and the modelling of vagueness using the theory of rough sets. In chapter 3 we present a literature survey focusing on the treatment of risk in each of the value function, outranking, and metric MCDM schools. We emphasise in particular the use of a small number of state scenarios as a treatment of risk, arguing that the resulting scenario-based approaches offer an integrated view of risk modelling within the MCDM framework, and facilitate a deeper exploration of the set of uncertain outcomes by the decision maker. We conclude chapter 3 and the literature survey with a summary of important research questions.

We then select three of those research questions for closer investigation in the remainder of the thesis, using a simulation model in the context of the value function methods. The questions considered are:

1. Can the 'ignoring' of uncertainty by simplification strategies be justified?

2. Do non-idealities known to occur in practice significantly affect the results under conditions of risk?
3. To what extent do results differ depending on how many scenarios are chosen and what scenario selection policies are used?

In chapter 4 we describe the structure and implementation of the simulated value function approach. We consider the simulation of an additive utility model taking into account the full range of uncertain outcomes, two models based on using expected attribute values as inputs to an additive utility model, and four scenario-based models differing only in the way that the scenarios are selected. The relative success of each model can be investigated by comparison to the results achieved by a multiplicative utility model based on an idealised preference structure. The chapter concludes by briefly proposing how similar investigations may be conducted in the context of each of the outranking and metric methodologies.

The results of simulating different value function methods are presented in three experiments making up Chapter 5. The experiments are demarcated in order to focus in the first experiment purely on the use of scenarios in value function modelling, and then investigate the interaction between the use of scenarios and certain practical non-idealities introduced in the latter two experiments. The second experiment investigates the effects of errors in the assessment of criterion weights and scenario relative likelihoods, while the final experiment investigates the effects of criterion omissions, independence violations, shifts in reference levels and piecewise modelling of marginal utility functions. The chapter concludes with a *post hoc* evaluation of the simulation structure presented in chapter 4.

The final chapter is dedicated to the identification of conclusions and recommendations for future research. The key conclusions from among those discussed are:

1. the danger inherent in using a scenario-based model has been understated in the literature, and results can be poor if the approach is applied uncritically. In particular, the notion of diversity as a basis for scenario selection needs to be more closely investigated for scenario planning and MCDM to be more meaningfully integrated.
2. a simplification of the modelled *preference* structure appears to have less severe consequences for the integrity of results than a simplification of the modelled *stochastic* structure.
3. the use of a model employing expected attribute values can produce good results if the expected attribute values are closely approximated. However, the danger exists that if the expected values are not closely approximated, the results are extremely poor.

Contents

1	Introduction	1
1.1	Background to the Problem	1
1.2	Statement of the Problem	2
1.3	Objectives of the Study	2
1.4	Limitations of the Study	3
1.5	Plan of Development of the Thesis	3
1.6	Terminology and Preliminary Notation	4
2	Introducing Uncertainty into MCDM	6
2.1	Introduction to Deterministic MCDM	6
2.1.1	Value Function Methods	6
2.1.2	Outranking Methods	8
2.1.3	Metric Methods	12
2.2	Aspects of Uncertainty	14
2.2.1	Zimmerman's Taxonomy of Uncertainty	15
2.2.2	Klir and Folger's Taxonomy of Uncertainty	16
2.2.3	French's Taxonomy of Uncertainty	17
2.3	Integrating Uncertainty into MCDM	18
2.3.1	Risk	18
2.3.2	Imprecision	19
2.3.3	Vagueness	20
2.4	Uncertainty in MCDM: A Morphological Box	20
2.5	A Note on Fuzzy Sets	22
2.6	A Note on Rough Sets	25
3	External Uncertainty: The Treatment of Risk	29
3.1	Value Function Methods	31
3.1.1	Multiple Attribute Utility Theory	31
3.1.2	Scenario-based Approaches	36
3.1.3	Non-Expected Utility Theory	41
3.2	Outranking Methods	44
3.2.1	Pairwise Reduction of Probability Distributions	44
3.2.2	Construction of a Distributive Outranking Relation	47
3.2.3	Scenario-based Approaches	48

3.3	Metric Methods	52
3.3.1	Chance-constrained Goal Programming	54
3.3.2	Other Stochastic GP Models	57
3.3.3	Scenario-based Approaches	58
3.4	Explicit Risk Attributes	63
3.5	Summary of Research Questions	66
4	Simulation Studies in MCDM	69
4.1	Research Experiments in MCDM	69
4.2	The General Simulation Structure	70
4.3	Implementation of a Simulated Value Function Method	71
4.3.1	The Problem Context	71
4.3.2	The Idealised Preference Structure	73
4.3.3	The Modelled Non-Idealities	75
4.3.4	Treatment of Uncertain Outcomes	77
4.3.5	Comparisons of Rank Orders	80
4.4	Parameters of the Simulation Experiment	81
4.5	Proposed Implementation of a Simulated Metric Method	83
4.5.1	Initial Aspirations	84
4.5.2	Computation of Deviations	85
4.5.3	Aggregation of Deviations	86
4.5.4	Revision of Aspirations	87
4.6	Proposed Implementation of a Simulated Outranking Method	88
4.6.1	Concordance and Discordance	89
4.6.2	Aggregation and Exploitation	90
5	Simulation of Value Function Methods	92
5.1	Results of Simulation Experiment 1	92
5.1.1	ANOVA Results	93
5.1.2	Characterisation of Effects	93
5.1.3	Summarised Inferences	104
5.2	Results of Simulation Experiment 2	107
5.2.1	ANOVA Results	107
5.2.2	Characterisation of Effects	108
5.2.3	Summarised Inferences	113
5.3	Results of Simulation Experiment 3	116
5.3.1	ANOVA Results	116
5.3.2	Characterisation of Effects	117
5.3.3	Summarised Inferences	125
5.4	Post-hoc Reflections on the Simulations	129

6	Conclusions and Future Research	131
6.1	Insights from the Simulation Study	131
6.2	Insights from the Literature Survey	135
6.3	Themes for Future Research	138
6.4	A Final Comment	141

University of Cape Town

List of Figures

2.1 A Graphical View of Uncertainty Modelling	15
---	----

University of Cape Town

List of Tables

4.1	Simulated levels of attribute variability	81
4.2	Parameters of the four simulated preference structures	82
5.1	Parameters of experiment 1	92
5.2	Grand average performance measures for experiment 1	94
5.3	Effect of reference level (τ_j)	97
5.4	Effect of utility at reference level (λ_j)	97
5.5	Effect of interaction between reference level (τ_j) and utility at the reference level (λ_j)	98
5.6	Effect of attribute variability (B)	99
5.7	Effect of interaction between reference level (τ_j) and attribute variability (B)	99
5.8	Effect of multivariate risk proneness (κ)	101
5.9	Effect of interaction between reference level (τ_j) and multivariate risk proneness (κ)	101
5.10	Effect of the number of selected scenarios (s)	102
5.11	Probabilities that the true best alternative is placed first or second in the model rank order	103
5.12	ANOVA results for POSN scores in experiment 1	105
5.13	ANOVA results for RANK scores in experiment 1	106
5.14	Parameters for experiment 2	107
5.15	Average POSN and RANK scores for experiment 2 (error) and corresponding experiment 1 (no error) problem contexts	108
5.16	Average rank correlations for experiment 2 (error) and corresponding experiment 1 (no error) problem contexts	108
5.17	Effect of errors in the assessment of criterion weights (e_w)	110
5.18	Effect of interaction between attribute variability (B) and weight assessment errors (e_w)	111
5.19	Effect of interaction between the number of selected scenarios (s) and weight assessment errors (e_w)	111
5.20	Effect of errors in the assessment of scenario probabilities (e_{Pr})	112
5.21	Probabilities that the true best alternative is placed first or second in the model rank order	112
5.22	ANOVA results for POSN scores in experiment 2	114

5.23	ANOVA results for RANK scores in experiment 2	115
5.24	Parameters for experiment 3	116
5.25	Average POSN and RANK scores for experiment 3 (non-ideal) and corresponding experiment 1 (ideal) problem contexts	117
5.26	Average rank correlations for experiment 3 (non-ideal) and correspond- ing experiment 1 (ideal) problem contexts	118
5.27	Effect of number of segments modelling the partial utility functions (v)	120
5.28	Effect of reference level shifts (e_r)	121
5.29	Effect of interaction between the number of segments (v) and reference level shifts (e_r)	121
5.30	Effect of violating independence (γ)	122
5.31	Effect of omitting criteria (o)	123
5.32	Effect of interaction between the number of selected scenarios (s) and omission of criteria (o)	124
5.33	Probabilities that the true best alternative is placed first or second in the model rank order	124
5.34	ANOVA results for POSN scores in experiment 3	127
5.35	ANOVA results for RANK scores in experiment 3	128

Chapter 1

Introduction

1.1 Background to the Problem

The nature of human decision making has over time attracted and fascinated a far wider group than the mathematicians, operations researchers, psychologists and economists which today make up most of the field of multiple criteria decision making (MCDM). The questions of how we do and, what is usually a different matter, should make decisions are of sufficient scale and scope to provide a home to a multitude of disciplines.

Of course much of human decision making must be made under conditions of uncertainty. Although as we shall see later the notion of uncertainty is often used as a loosely defined umbrella-term for a number of more specific concepts, intuitively we recognise it as an epistemic condition indicating a lack of accuracy, confidence or precision. Inasmuch as reality can be considered an infinitely detailed construct incapable of being fully comprehended by humans (for example, [105]), all decisions can be considered to be subject to conditions of uncertainty. The task of those developing and applying MCDM models is therefore 'to look for precision in each class of things just as far as the nature of the subject admits' [91].

The freedom of interpretation permitted by the previous quote is evident in the diverse development of MCDM models. Nevertheless it is also a frequent cause of concern that the different emphases implicit in various methodologies give rise to different results. The nature of MCDM is that it provides support and structure for exploration rather than hard solutions to the problem, and as such the different solutions can be viewed as realisations of the different aspects explored and the different languages of exploration. However, it is important that those aspects of the problem emphasised by the various methodologies are explicitly known and appreciated, and in particular which aspects of the decision problem each methodology is most sensitive to. This has sometimes been missing from methodological developments in MCDM.

1.2 Statement of the Problem

Within the field of MCDM, a specific form of decision is specified that is nevertheless sufficiently general to encompass a wide range of problems. The problem context involves a decision maker (DM) who must consider a choice of action given a number of alternatives. This choice is governed by a set of objectives, usually conflicting in any non-trivial problem, that are used as means for evaluating the alternatives. The resolution of this conflict through the progression to a final choice lies at the core of MCDM, but encompasses many aspects. Firstly, there is the considerable problem of identifying the appropriate alternatives and objectives, or structuring the problem. Then, we need to construct a suitable model not only to provide a solution to the problem, but also to facilitate the DM gaining meaningful insight into the decision context, the choices available, and his or her own preferences. Lastly, the information and insight obtained using the model needs to be interpreted and implemented in a meaningful way that serves the original aims of the decision analysis.

The construction of a model depends on the evaluation of how the alternatives perform on the various objectives, so that these performances may be weighed up against each other in the progression to a resolution. The problem therefore becomes considerably more difficult when those evaluations cannot be specified with precision; that is, they are uncertain. In such a case not only must we construct a decision model, but we need to accommodate the uncertain evaluations in some manner. Moreover, the two aspects are not independent of each other.

The overarching problem can therefore be simply stated: given that the performances of the alternatives are not known with certainty, how should we go about constructing a decision model?

1.3 Objectives of the Study

There are four main objectives which are the subject of this study

1. To investigate what aspects of uncertainty are relevant to MCDM.
2. To investigate how the uncertainties identified as relevant are treated by various different MCDM methodologies.
3. To investigate interrelationships between the modelling of the uncertainty aspects and the multiple criteria aspects of the decision problem.
4. To highlight ways in which the methodologies identified by objective 2 might be improved based on the implications of objective 3.

1.4 Limitations of the Study

- In general we consider that the aim of the decision analysis is to place the alternatives in a (not necessarily complete) rank order, that is we assume the ranking problematique [75]. We therefore exclude from consideration several alternate problematiques (see [7]), although the methods are in general similar and would require only minor modifications.
- We adopt the conventional but often-criticised assumption that the problem context has been structured so that the starting point is a well-defined decision table. We therefore ignore important issues such as defining sets of alternatives and criteria, and isolating the decision problem from its broader environment. The full decision process is described in, for example, [7] and [75].
- Although we initially identify three categories of uncertainty relevant to MCDM, in the interests of time and resource savings we pursue in detail only uncertainty relating to the physical randomness of future events. Investigations into the other two types of uncertainty are limited to brief discussions of the key issues in each.
- We also initially present the three MCDM methodologies of value function, out-ranking and metric methods. A simulated approach to decision making under uncertainty was only performed for the value function methods. The value function methods were chosen on the basis that their axiomatic foundations allowed a simulation model to be constructed relatively quickly and easily, based on extending existing work done by Stewart [86]. Proposed simulation approaches and key issues are presented for the other methodologies. It is hoped that these approaches will be implemented in the future.
- A final limitation is the reliance on simulation approaches to achieve objectives 3 and 4. Although many of the issues are best suited to behavioural studies, this is neither in the time or cost framework of the current study. This obviously influences the aspects chosen for closer investigation; we postpone a more detailed discussion until chapter 4.

1.5 Plan of Development of the Thesis

Chapter 2 presents the three dominant MCDM methodologies from a deterministic context. The deterministic framework is both conceptually easier and more developed than its uncertain counterpart, and so provides a natural starting point for any further investigation. Then, we structure a formal view on the uncertainty that we wish to consider further, providing a working definition and identifying three categories of uncertainty that are relevant to MCDM. A morphological framework for decision making under uncertainty is provided before we briefly consider the issues around two

of these categories, leaving the third to a more detailed investigation in chapter 3.

Chapter 3 presents the current state-of-the-art in the treatment of decision problems where the performances of the alternatives are dependent on unknown future events. The chapter concludes with a review of those issues identified as significant during the chapter, and attempts to structure them into a suitable framework for subsequent solution.

Chapter 4 begins with the selection of the key research questions to be considered in the remainder of the thesis. A brief introduction to simulation in MCDM is then provided, followed by an explanation as to why this mode of investigation was chosen for this study. It then describes in detail the simulation approach used in the remainder of the thesis to study the value function method, and provides two proposed simulation approaches that may be useful in future studies of outranking and metric methods.

Chapter 5 presents the results of the simulation approach outlined in chapter 4, following which the final chapter makes conclusions and recommendations by drawing together the results of the simulation study and important aspects arising from the literature survey.

1.6 Terminology and Preliminary Notation

We make extensive use of the following basic MCDM features:

Alternative: Implicit in any decision is a choice between various courses of action. The set of alternatives is made up of the possible courses of action facing the DM. Usually we refer to a discrete set of explicitly identified alternatives, although in some circumstances the alternatives are defined implicitly in terms of a set of constraints.

Criterion: A criterion is a basis or dimension on which the alternatives may be compared and judged. We refer without loss of generality to criteria in the increasing sense only i.e. more is always preferred.

Attribute: Often criteria may be fairly esoteric notions (comfort, employee satisfaction). An attribute is a well-defined measure of performance associated with each criterion that may be used to unambiguously rank the alternatives according to the aspect identified by the relevant criterion.

Decision Variable: When the alternatives are implicitly defined by constraints, they are expressed in terms of a number of unknown variables for which the

DM must make a decision. Combinations of these decision variables can therefore be considered to be alternatives from which the DM must decide the most favourable option.

Objective: Objectives, like attributes, are measures of performance that may be used to judge performance. However, they are continuous analogues of attributes, defined in terms of decision variables.

We refer to a set of n alternatives indexed by $i \in \{1, \dots, n\}$, and a set of m criteria indexed by $j \in \{1, \dots, m\}$. Each criterion is associated with an underlying attribute, so that by z_{ij} we denote the evaluation of alternative i on criterion j . It is often useful to think of these evaluations in the framework of an $n \times m$ ‘decision table’ \mathbf{z} . When referring to specific alternatives in the full set, we typically refer to alternatives a and b , and denote their evaluations on criterion j as $z_j(a)$ and $z_j(b)$. In the continuous framework, the set of r decision variables is represented by $\{x_1, \dots, x_r\}$, while objectives are referenced in the same way as criteria, using $j \in \{1, \dots, m\}$. The z_{ij} will in this case be represented as a linear weighted sum of the decision variables.

The z_{ij} may be deterministic or stochastic. Where they are stochastic, we associate with each random variable z_{ij} a probability density function $f_{ij}(z_{ij})$ and a cumulative probability distribution $F_{ij}(z_{ij})$, which we abbreviate as f_{ij} and F_{ij} respectively. When we refer to specific alternative a on criterion j we break from the convention above and denote the probability distributions f_{aj} and F_{aj} .

Scenario: We consider the z_{ij} as random variables, so that there is in principal a single multivariate probability distribution governing the joint realisation of all the z_{ij} . Each realisation of this multidimensional random variable corresponds to a potential future state of the world, so that the underlying form of the distribution is likely to be extremely complex and unattainable from a practical perspective. We therefore make no attempt to specify the full multivariate probability distribution, but rather seek only to *characterise* it using a small number of potential realisations or states, which we term the scenarios. The term ‘scenario’ is thus used consistently with the scenario planning literature (e.g. [96]) to refer to an internally consistent future state of the world, although we extend this definition to include a precise probabilistic interpretation. Although we do not make an explicit assumption about the characteristics of these scenarios, they will in general be constructed either as (incomplete) ‘snap-shots’ of a future state, or a plausible evolution from the present state into the future.

We refer to the set of scenarios using the index k , $k \in \{1, \dots, p\}$. We then write the consequences of scenario k for alternative i on criterion j as z_{ijk} . When referring to a specific alternative a the notation becomes $z_{jk}(a)$ as before.

Chapter 2

Introducing Uncertainty into MCDM

2.1 Introduction to Deterministic MCDM

Let us assume that the decision process has progressed through the problem structuring phase to arrive at a set of n alternatives for closer evaluation, as well as a set of m criteria. Each of these criteria is associated with an underlying attribute, allowing for the evaluation of the performance of alternative i on criterion j to be represented by an attribute value z_{ij} . We assume in this section that we know precisely the z_{ij} , and assemble these evaluations in an $n \times m$ decision table, \mathbf{z} .

2.1.1 Value Function Methods

The value function methods aim to structure and subsequently exploit a function V that will associate with each alternative a real number such that for any two alternatives a and b ,

$$a \succ b \iff V(a) > V(b) \tag{2.1}$$

$$a \sim b \iff V(a) = V(b), \tag{2.2}$$

where $a \succ b$ means that a is preferred to b , and $a \sim b$ indicates indifference between a and b . Although assumptions of completeness and transitivity guarantee the existence of $V(\cdot)$, the simplification of this so-called value function is of paramount importance from the practical perspective of ease of use and understanding. An additive representation $V(a) = \sum_{j=1}^m w_j v_j(a)$, where $v_j(a)$ is an abbreviation of $v_j(z_j(a))$, is possible if preferences between two alternatives can be expressed without recourse to the level of the criteria for which evaluations are equal. This condition is termed preferential independence. Within the resulting marginal value functions, the framework of value tradeoffs is employed as a vehicle for strength of preference information. The question essentially asked is ‘How much of attribute 1 is required to compensate for a fixed

decrease in attribute 2?'. It is this tradeoff issue that is explicitly formalised in the value function methods.

Intra-criterion Information

The elicitation of value scores is a constructive exercise in elucidating often murky thought patterns and value judgements from the DM, and is a cornerstone of the learning aspect of the value function methods. Four broad elicitation approaches are possible:

1. The explicit definition of a marginal value function for each attribute, either over the set of real alternatives or a broader hypothetical set encompassing, for example, concepts of best- and worst-possible performance.
2. The construction of an underlying qualitative value scale, examples of which are the Likert and Beaufort scales, again over either a real or hypothetical set of alternatives.
3. A direct rating of the value of each alternative, usually by referring to performance relative to the best- and worst-performing alternatives.
4. By inference from a set of pairwise strength of preference evaluations, as in the Analytic Hierarchy Process (AHP) developed by Saaty [78].

Generally, the values on each criterion are standardised, for example by assigning a value of 0 to the worst alternative (real or hypothetical) and 100 to the best. The additive form of the value function implies an interval scaling for the attribute values. The selection of a zero point on the value scale is in many cases arbitrary, being set at either the worst performance among a set of alternatives or a perceived worst-case. As such it is only the ratios of differences between attribute values that have an absolute meaning. The choice of elicitation approach obviously has enormous practical implications for both the elicitation process and the progression of the analysis, well beyond the scope of this introduction; these issues are dealt with in [7].

Inter-criterion Information

When constructing value scores, the performance evaluations took place within each criterion separately. Naturally not all criterion can be considered equal – some may be more important than others, so that a weight may be assigned to each criterion. Although the precise meaning of ‘importance’ is often difficult to pin down, the use of value function methods implies a very specific interpretation for the criterion weights. Essentially the weights are scaling constants indicating the relative value of swings between two standardised points on the value function, for example a swing from worst to best performance. The decision maker is required to identify the criteria in which a swing from worst to best is most highly valued, and then judge the value of similar 0 – 100 swings on the other criteria, using the favoured criterion as a reference.

This is known as the swing-weight interpretation, and clearly encompasses the range of the attribute values on a particular criterion as well as more esoteric notions of ‘importance’.

Aggregation and Exploitation

The global value scores are easily obtained from the additive form once the scores and weights have been computed, so that the global rank ordering of the alternatives follows trivially. These results are advocated as a mechanism for further discussion and learning, as well as possible restructuring where the various issues surrounding the sensitivity analysis suggest that this is appropriate.

Why Use a Value Function Approach?

The use of a value function approach does require compliance with some fairly strong behavioural assumptions, at least in principle. Where the DM is comfortable with these constraints, no other MCDM approach provides as structured a framework for the investigation of preferences and tradeoffs. This structure, in particular the formality and simplicity of the criterion weights and additive aggregation, is likely to be invaluable in facilitating a deeper understanding of the problem. Essentially the model is simple enough to be considered to ‘run in the background’ while freeing the DM to consider only the tradeoff issues at hand. Unfortunately it is this simplification which is the biggest potential pitfall of the value function methods. Where the DM is not comfortable with the underlying assumptions, but nevertheless uses the methods, a decision may be ‘forced’ out of the DM. In such cases there is an inevitable question as to whether ‘the final decision is justified by the characteristics of the problem or by the mathematical properties of the method’ [98]. The value function approach is therefore ideal for direct use with the DM in the interactive environment of a workshop, with the strong proviso that the DM is comfortable with the behavioural assumptions implied by the model.

2.1.2 Outranking Methods

The outranking methods aim at constructing a preference relation over the set of alternatives such that alternative a outranks alternative b if there are sufficient arguments to support the assertion that ‘ a is at least as good as b ’, and no arguments of sufficient strength to refute this assertion. The aim is intentionally rather vague, which allows several interpretations as to what constitutes sufficient supporting and negating arguments. The outranking relation is applied directly to the underlying attribute values, without recourse to an intermediary such as a value function. This releases the outranking approach from the requirements of completeness and transitivity. This is a fundamental feature of the outranking methods: it is maintained that MCDM is a constructive process that should not impose on a DM the demand of decisiveness. An additional possible outcome of the decision analysis is therefore incomparability,

which covers the situation where a lack of information or inadequately-formed preferences preclude either strict preference or indifference. Any further characterisation of the outranking relation, particularly axiomatically, is a difficult and ongoing research problem [9, 98].

Intra-criterion Information

The construction of the outranking relation relies on the formalisation of the supporting and negating arguments discussed earlier. Although the different formalisations give rise to several different outranking methods, two common features are indices measuring concordance and discordance. The concordance index $c_j(a, b)$ builds up the evidence supporting the statement that a outranks b on criterion j , while the discordance index $d_j(a, b)$ provides evidence against the assertion. Discordance takes the form of a veto, so that if a performs far worse than b on even one criterion, then one may not say that a outranks b , regardless of the information on the other criteria. Discordance lends a strongly non-compensatory element to the outranking methods – good performance on one criterion does not necessarily trade off bad performance on another criterion. It requires the specification of a veto threshold t_j^v by the DM, representing the smallest interval $[z_j(a), z_j(b)]$, where $z_j(a) \leq z_j(b)$, which is sufficient to refuse the assertion that a outranks b . We define the criterion-wise concordance and discordance indices for two outranking methods that will be referred to again later in the study.

- **ELECTRE I:** A criterion j is included in the aggregation of the global concordance measure if $z_j(a) \geq z_j(b)$, that is

$$c_j(a, b) = \begin{cases} 1 & \text{if } z_j(a) \geq z_j(b), \\ 0 & \text{otherwise.} \end{cases} \quad (2.3)$$

A criterion j is considered to be discordant if $z_j(a) + t_j^v \leq z_j(b)$, that is

$$d_j(a, b) = \begin{cases} 1 & \text{if } z_j(a) + t_j^v \leq z_j(b), \\ 0 & \text{otherwise.} \end{cases} \quad (2.4)$$

These specifications are also used in the **ELECTRE II** method.

- **ELECTRE III:** The concordance and discordance measures are extended into a fuzzy environment by the introduction of two DM-specified thresholds in addition to the veto threshold t_j^v defined earlier: an indifference threshold t_j^i and a strong preference threshold t_j^p , which also may depend on $z_j(a)$. Concordance is then calculated as

$$c_j(a, b) = \begin{cases} 1 & \text{if } z_j(a) + t_j^i \geq z_j(b), \\ \frac{z_j(b) - (z_j(a) + t_j^p)}{t_j^i - t_j^p} & \text{if } z_j(a) + t_j^i \leq z_j(b) \leq z_j(a) + t_j^p, \\ 0 & \text{if } z_j(b) \geq z_j(a) + t_j^p. \end{cases} \quad (2.5)$$

Similarly the discordance measure $d_j(a, b)$ is given by

$$d_j(a, b) = \begin{cases} 0 & \text{if } z_j(a) + t_j^p \geq z_j(b), \\ \frac{z_j(b) - (z_j(a) + t_j^p)}{t_j^v - t_j^p} & \text{if } z_j(a) + t_j^p \leq z_j(b) \leq z_j(a) + t_j^v, \\ 1 & \text{if } z_j(b) \geq z_j(a) + t_j^v. \end{cases} \quad (2.6)$$

Inter-criterion Information

The aggregation of the criterion-wise measures of concordance and discordance into corresponding global measures naturally requires some form of weighting procedure. However, the non-compensatory aspect of the outranking methods means that the swing-weight interpretation is inappropriate. The weight on each criterion is viewed as ‘the number of votes’ allocated to that criterion in the vote to test the assertion that alternative a outranks b . This is most clearly seen in the context of a ‘pure’ non-compensatory approach such as ELECTRE I. For the more sophisticated outranking methods, the weights are partly numbers of votes and partly tradeoffs, in the regions where the methods are non-compensatory and compensatory respectively (see [98] for further details). In the absence of definitive operational and behavioural interpretations the criterion weights may end up being judged on notions of ‘importance’ that are known to be fairly arbitrary. We again examine the two outranking methods.

- ELECTRE I: The global concordance measure is computed as

$$C(a, b) = \frac{\sum_{j=1}^m w_j c_j(a, b)}{\sum_{j=1}^m w_j} \quad (2.7)$$

The global discordance measure is given by

$$D(a, b) = \begin{cases} 0 & \text{if } \exists j : d_j(a, b) = 1 \\ 1 & \text{otherwise} \end{cases} \quad (2.8)$$

- ELECTRE III: The global concordance measure is defined as for ELECTRE I, while the global discordance measure is given by

$$D(a, b) = \prod_{j=1}^m f(d_j(a, b), C(a, b)) \quad (2.9)$$

where

$$f(d_j(a, b), C(a, b)) = \begin{cases} 1 & \text{if } d_j(a, b) < C(a, b), \\ \frac{1 - d_j(a, b)}{1 - C(a, b)} & \text{if } d_j(a, b) > C(a, b) \end{cases} \quad (2.10)$$

Aggregation and Exploitation

The construction of the outranking relation involves the synthesis of information regarding the concordant and discordant pieces of information relating to the assertion that alternative a outranks alternative b . Again, the exact manner in which this synthesis occurs depends on the outranking method used.

- ELECTRE I: Alternative a is said to outrank alternative b if $C(a, b)D(a, b) \geq t^C$, where t^C is a user-specified concordance level indicating the ‘majority vote’ required to pass the assertion.
- ELECTRE III: No firm conclusion is drawn from the concordance and discordance results; rather a ‘credibility index’ $S(a, b)$ is attached to the assertion that a outranks b , where $S(a, b) = C(a, b)D(a, b)$. The credibility index takes on values between 0 and 1, increasing with the strength of the evidence supporting the assertion: it has therefore been compared with the notion of a membership function in fuzzy set theory [7].

The relaxation of the assumptions of completeness and transitivity means that preferences may not be represented by a numerical function, so that the ranking of the alternatives becomes a non-trivial procedure for which more than one method exists. Generally though, two ‘distillations’ are produced. One is a descending rank order, which begins with the alternatives that outrank most alternatives and ends with alternatives that outrank very few or no alternatives. The other is an ascending rank order, starting with alternatives that are outranked by many other alternatives and building upward to alternatives that are outranked by few or no alternatives. These two rank orders give different information and decision support, and should at least initially be considered separately. Depending on the divergence of the rank orders it may be possible to meaningfully synthesise them using an intersection operation. Further details may be found in [98]

Why Use an Outranking Approach?

The outranking methods aspire to lofty ideals of dynamicism and constructivism in the decision process, but an essential question is whether they are able to deliver on these ideals. The notion of incomparability permits an indecisiveness which is a natural part of decision making, yet the methods demand a high level of comprehension in the specification of behaviourally ill-defined thresholds and weights. The unstructured nature of the approach (in terms of the consequences of the lack of axiomatic foundations) is a source of potential confusion among facilitators and decision makers, yet it also allows the decision-aiding process to go on when other methods may fail; for instance when compensations between criteria are unclear or when underlying units of criteria are very heterogenous. The use of tradeoffs may be considered undesirable in certain politically-charged environments, although any MCDM method must necessarily bring to light some value judgements that might be

considered controversial. Nevertheless, the outranking approaches seem ideally suited to high-level decision making in purposefully ill-defined and messy environments. In such environments, flexibility may be considered an asset above all others.

2.1.3 Metric Methods

Goal programming is the oldest of the MCDM approaches considered here, dating back to work done in the 1950's by Charnes and Cooper [15]. The underlying cognitive model is that the DM sets goals or aspiration levels for the objectives under consideration, and then evaluates prospective alternatives via a dynamic and iterative comparison with the aspiration levels. It is the explication of this comparison that gives rise to the many varieties of goal programming which we have grouped under the umbrella of metric methods. In 1976, Simon [80] proposed a descriptive model of decision making that became a motivation for the use of goal programming. His notion of 'satisficing' supposed that a DM will search for an alternative offering satisfactory performance on all criteria without necessarily attempting to maximise this performance. In the considered MCDM context, this translates to the DM seeking to achieve a satisfactory level of performance on the most important criterion before considering other criteria. In what follows, criteria have been defined so that they are satisfied when their goals are exceeded i.e. in a maximising sense. Accommodations for other forms of deviations are easily made, but the conformity simplifies the presentation.

Intra-criterion Information

The comparison of an aspiration to the performance of an alternative naturally takes the form of a distance measure, so that we attempt to minimise the underachievement i.e. positive deviation, between the aspiration g_j and the performance of an alternative $z_j(a)$ on a particular criterion j . This distance-based measure of performance implies that attributes should be defined over cardinal and not categorical scales. The aspirations themselves may be specified along a spectrum ranging from highly optimistic, even unachievable, values, to strict lower bounds on acceptable performance. Placing a set of aspirations in this spectrum may be difficult for even an experienced DM, and the cognitive processes guiding aspiration selection are poorly understood. Belton and Stewart [7] have suggested that 'moderately optimistic' aspirations be set, so that aspirations on certain subsets of criteria are feasible, but no alternative satisfies all aspirations simultaneously.

Inter-criterion Information

The progression towards a decision requires a comparison of the deviations in different criteria, introducing a need for some weight measuring the relative importance of each criterion. The weights are to be interpreted as for the value function methods, in a swing-weighting sense. Notions of relative importance are integrated with the different

ranges on each criterion, so that the weights take the form $w_j = I_j/R_j$, where I_j is a measure of the relative importance of the swing from worst to best values, and R_j is the range of values for criterion j .

Aggregation and Exploitation

The global comparison of alternatives requires the aggregation of the single criterion deviations into a measure of the distance between the performance of an alternative and the aspirational performance. A distance computed over such a multi-dimensional criterion space implies that a choice of some metric is necessary. The general goal programming formulation is

$$\min \left[\sum_{j=1}^m [w_j \delta_j]^\alpha \right]^{\frac{1}{\alpha}} \quad (2.11)$$

where w_j is the weight assigned to the deviation δ_j from goal g_j , for each criterion j , and α denotes the choice of metric. Typical choices for α are one, giving the Archimedean goal programming formulation [7], and infinity, which describes the Tchebycheff formulation. The choice of metric implies quite different behavioural models: the Archimedean formulation minimises a weighted sum of deviations, while the Tchebycheff formulation minimises the maximum weighted deviation. The Tchebycheff approach therefore seems more closely aligned with the interpretation of satisficing introduced earlier.

A rank order can be established on the basis of the above minimisation. However, the sensitivity of results to the choice of goals and the difficulty of selecting well-balanced goals means that this rank order should be used as a basis for further exploration of tradeoffs rather than for alternative selection. This has given rise to a number of so-called interactive procedures to guide the iterative reassessment of aspirations in order to comprehensively search the decision space (for example, see [89]).

Why Use a Metric Approach?

The metric approaches generally require less preference information than either value function or outranking methods. The methods are fairly simple conceptually, and model the intuitively appealing heuristic of satisficing for decision making. There is little in the methods to confuse even an inexperienced DM; an aspiration level is an intuitively meaningful concept even if its influences in terms of contextual and psychological factors are not well understood. Whether this lack of depth in the preference information is seen as an advantage or a disadvantage is likely to depend on the context of the decision problem. In some circumstances, it may be impossible or unnecessary to make the detailed preference judgements needed to construct a value function or outranking relation. Typical situations might include the early stages of a decision process, where the focus is on the selection of a set of acceptable

alternatives for deeper analysis, or routine decisions, where the difference between an acceptable alternative and an optimal one is unlikely to be sufficient to warrant a more detailed analysis. In other circumstances, it may be more important for the DM to gain insights into the nature of his or her preferences through the tradeoffs that are available. A metric method is unlikely to be as facilitative in these respects as a value function or outranking method. In particular, metric methods should not be used to avoid asking the DM difficult questions.

2.2 Aspects of Uncertainty

The precise characterisation of uncertainty is notoriously problematic, being based more often than not on the background of the researcher (computer science, systems engineering, decision analysis, ...) and the generally closely related context of the problem. In fact, even a broad definition of uncertainty is lacking from many texts dealing expertly with its incorporation and reduction [50, 28]. These texts prefer to provide a taxonomy of uncertainty in a more precise manner, leaving the actual broader concept of uncertainty as a cloud hanging over the classifications.

Although the depth of an information-theoretical treatment of uncertainty is well beyond the scope of this thesis, it is certainly possible and beneficial to at least have in some sense a firm notion of the uncertainty we are trying to model. Following Zimmerman [115], the notion of uncertainty is founded on two conjectures

1. Uncertainty is viewed as a 'situational property'; that is, it depends heavily on the context of the problem under study.
2. Uncertainty is a subjective impression which is as such intertwined with the nature and behaviour of the decision maker. Most importantly, this implies a dependency on the quality and quantity of information available to the DM.

The view of uncertainty assumed in this thesis can then be represented graphically by adapting the framework provided by Zimmerman to a multicriteria context, as shown in figure 2.1. The uncertainty model is therefore advocated as a 'filtering' of the information regarding the uncertain phenomenon prior to the perception of this information by the DM. The DM does not view the uncertain phenomenon directly, but rather uses the uncertainty model as 'glasses through which to view the uncertain situation' [115]. As such, the selected uncertainty model should be compatible with both the features of the uncertain phenomenon and the decision maker. As the graphical representation suggests, the selection of an uncertainty model is influenced by and influences the choice of an MCDM model, with this dual selection being dependent on both the problem and DM context. The implication is that no single uncertainty theory and consequent model may claim to be universally superior, and that an informed, integrated view of uncertainty modelling is necessary.

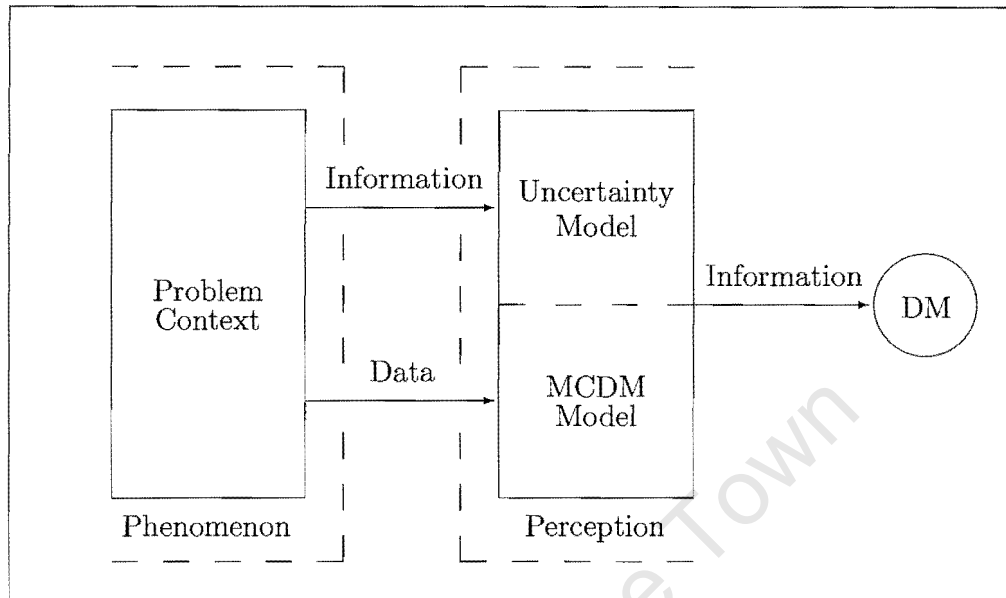


Figure 2.1: A Graphical View of Uncertainty Modelling

Having provided a notion of uncertainty, the primary objective is to break it down into more practical and pragmatic segments by providing a rough taxonomy of uncertainty properties and, by way of the previous argument, of the uncertainty models themselves. The nature of this taxonomy is neither exhaustive nor agreed upon, although generally there is at least some common ground. In the following section three broadly representative taxonomies are discussed. It should be emphasised that the taxonomies are not examined from the perspective of whether or not it is practically meaningful to model the types of uncertainty in the context of MCDM, but rather from a purely classificatory standpoint.

2.2.1 Zimmerman's Taxonomy of Uncertainty

Uncertainty is represented by a four-valued vector characterised by the following properties.

1. The causes of uncertainty.
 - Lack of information
 - Abundance of information
 - Conflicting evidence
 - Ambiguity (complexity)
 - Measurement

- Belief
2. The quantity and quality of information.
 3. The type of information processing required by the DM.
 4. The language required as an output to the DM.

Aspects 2, 3 and 4 are not considered directly by this thesis, although they will naturally influence directly both the choice of the MCDM model and the nature of the decision process. Lack of information pertains to a situation in which the information is not sufficient, either qualitatively or quantitatively, to describe the problem deterministically. Two typical insufficiencies are information about the occurrence of a future state of nature, and information exactly describing objects i.e. approximation. The transition to certainty can only be achieved by the acquisition of more or better information. In contrast, the uncertainty relating to abundance of information pertains to the inability of human beings to absorb large amounts of data, and can only be reduced by transforming the data to more concise, meaningful information. Uncertainty due to conflicting evidence is observed whenever information points to more than one possible behaviour. Within the context of MCDM, it is conflicting evidence, and the resulting uncertainty regarding this conflict, which motivates the consideration of an MCDM model. The type of uncertainty called ambiguity stems from the different meanings of certain linguistic information i.e. good, bad. A key point is that the remedy may depend on collecting more information about the nature and context of such words, so that modelling ambiguity (for example, by way of fuzzy sets) may become secondary or even unnecessary. Measurement uncertainty relates to the inherent fallibility of all physical measurements, although within the context of MCDM it should be categorised with a lack of information. Finally, uncertainty of beliefs relates to the evolutionary nature of subjective information, and the lack of 'objectivity' in information. Naturally behind this type of uncertainty lies a wealth of debate regarding the state of objectivity of any kind in information. Further information about the taxonomy can be found in [115].

2.2.2 Klir and Folger's Taxonomy of Uncertainty

The following taxonomy is presented very much from the two authors' perspective of integrating information theory with fuzzy sets. As a result the omission of uncertainty arising from the unknown future state of the world is not overly surprising. More detail regarding the taxonomy is available in [50].

1. Vagueness
2. Ambiguity
 - Nonspecificity in Evidence
 - Confusion in Evidence

- Dissonance in Evidence

Vagueness relates to difficulties experienced in making precise distinctions between concepts or categories in the real world, and so is equivalent to the ambiguity of the Zimmerman taxonomy. In the Klir and Folger taxonomy, ambiguity pertains to 'situations in which the choice between two or more alternatives is left unspecified', the mathematical analogy of which is the one-to-many relation. Ambiguity is further divided into three classifications. Nonspecificity of evidence refers to the sizes of the subsets that have been identified as potential locations of a particular object, such that a larger cardinality implies a higher degree of ambiguity. Confusion of evidence exists whenever the subsets potentially locating an object overlap either partly or not at all, so that the information pointing to the different locations manifests itself as uncertainty, specifically ambiguity. A related uncertainty is dissonance in evidence, which is exhibited whenever the potential locations of an object are disjoint. The conflict arising from situations of confusion or dissonance are analogous to the earlier definition of conflicting evidence.

2.2.3 French's Taxonomy of Uncertainty

The taxonomy is constructed with respect to three methodological steps involved in conducting an analysis. The first phase is termed a modelling phase, involving the construction of what we have referred to as the problem context. The emphasis is on representing the real world in a tractable yet realistic model. The second phase is an exploration of the models built up in the modelling phase. Finally, the third phase interprets the results of the explorations and translates them into meaningful guidance for the decision maker. Within the context of the graphical representation of uncertainty provided in 2.1, the modelling phase refers to an investigation of the system locating the uncertain phenomenon, so as to provide a summary (in the form of the problem context) which will bridge the divide between the real-world phenomenon and the perception of it. The exploration phase is firmly based on the interaction between the uncertainty and MCDM model within the assumed problem context. The interpretation phase refers to the directing of suitable information from the exploration phase to the decision maker. Further details on both the methodological and taxonomical aspects may be found in [28].

1. Uncertainties expressed during modelling
 - Uncertainty about what might happen
 - Uncertainty about meaning (ambiguity)
 - Uncertainty about related decisions
2. Uncertainties expressed during the exploration of the models
 - Uncertainty arising from physical randomness

- Uncertainty about the evolution of future beliefs
- Uncertainty about the subjectivity of beliefs and preferences
- Uncertainty about the accuracy of calculations

3. Uncertainty expressed during interpretation

- Uncertainty about the nature of the model i.e. normative or descriptive
- Uncertainty about the depth of the analysis

The taxonomy is introduced in the context of the complete modelling of a decision, so that a great deal of importance is placed on the uncertainties relating to the structural issues surrounding the decision problem. While acknowledging the great importance of accommodating these structural uncertainties where possible, the primary focus of this thesis is on the interaction between MCDM models and uncertainty models taking place in the exploration phase. In many instances, the only recourse to the structural uncertainty represented here by modelling and interpretative uncertainty will be through further restructuring. In these cases the integrated decision model cannot ease these uncertainties, since it operates only on the perceptive model that is provided as an input. No model may capture what is not perceived, so that whether all criteria have been identified or whether another decision is related strongly enough to warrant inclusion lies beyond the scope of the exploration phase.

2.3 Integrating Uncertainty into MCDM

The taxonomies of the previous section were presented from various points of view: Zimmerman's from the area of systems engineering, Klir and Folger's from the more mathematically-driven theory of information, and French's from the perspective of a pragmatic decision analyst. The broad concept of uncertainty is not even defined in the latter two, so that we cannot be sure that they suppose to measure exactly the same concept. Nevertheless, the aim of the previous section was not to judge or even compare directly the various classifications, but merely to present them for consideration. In this section, an attempt is made to synthesise those parts of the taxonomies that are most relevant to MCDM. In doing so, we hope to arrive at a generally acceptable classification of uncertainty that, though non-unique and non-exhaustive, is applicable to MCDM. Three broad categories of uncertainty are defined: risk, ambiguity, and vagueness.

2.3.1 Risk

In many cases, it becomes necessary to predict the future outcomes of alternatives in order to compare them. For example, in comparing two holidays, it might be necessary to consider the possible weather at the time of the holiday. Risk corresponds in the French taxonomy to the uncertainty arising from physical randomness, and to the

inherent lack of information about this uncertain future in the Zimmerman taxonomy. Klir and Folger do not consider risk in their taxonomy due to the emphasis on fuzzy set theory. It is generally accepted that true physical randomness should be modelled using probability theory [28], although the precise nature of these probabilities is often not agreed upon. Some authors [93] make a distinction between cases where the probabilities of occurrence are known and unknown, calling them decisions under risk and decisions under uncertainty respectively. This distinction may serve some purpose in the debate between objectivists and subjectivists, but from the perspective of practical MCDM the treatment of both cases is identical. We therefore consider all decisions involving random or unknown outcomes to be decisions under risk.

The diverse connotations of the word 'risk' mean that it is often confused in a specific context. In the context of decision analysis, risk is often interpreted in the sense of a preference for a lottery over the expected value of the lottery. More broadly, risk is associated with the occurrence of an unpleasant event. Within the field of risk analysis, risk is considered to be a function of both the likelihood and consequences of a catastrophic event (see, for example, [52]). Neither of these uses should be confused with the uncertainty relating to the physical randomness of the outcomes.

2.3.2 Imprecision

Within the decision process, it is a necessary step that the decision maker must provide preferential information, for example when judging the weight of each criterion. The accuracy of this information is in a sense arbitrary, and decided by the needs of the analysis and the confidence of the decision maker. Nevertheless, no decision maker could possibly say that the weight of criterion j is exactly 0.4888 and not 0.4889. In any case where subjective information about beliefs or preferences needs to be elicited, the limited discriminatory power of the human mind is likely to raise some doubt or uncertainty in the decision maker, and it is this uncertainty which we call imprecision. This imprecision extends easily into physical processes. For example, although a car's top speed may be specified by the manufacturer to the nearest kilometer per hour, we may only be confident that it lies within 5 km/h either side of this and certain that it lies within 15 km/h either side. Imprecision is likely to be highly specific both to the context of the problem and the nature of the decision maker.

Imprecision encompasses to a large extent uncertainty arising from different linguistic meanings, although strictly they do model different limitations. The concept of ambiguity in Zimmerman's taxonomy refers to an imprecise *concept* usually represented by a linguistic term, while the uncertainty discussed in the previous paragraph refers to an imprecise *judgement* which nevertheless may have a precise linguistic meaning. There is an important consequence of the distinction: linguistic imprecision may be

resolved by more detailed definitions of the concepts under scrutiny, so that a more efficient and constructive remedy may be found via a return to the restructuring phase. Ambiguity in the Zimmerman taxonomy, nonspecificity of evidence in the Klir and Folger taxonomy, and uncertainty about the subjectivity of beliefs and preferences in the French taxonomy all fall under the umbrella of imprecision.

2.3.3 Vagueness

The final type of uncertainty relates to a more basic and fundamental uncertainty than the previous two. The idealised real world is assumed to be constructed from infinitely detailed objects which to all intents and purposes cannot be fully perceived by human beings. In an attempt to deal with this uncertainty, each human being assembles a set of information which is considered an adequate representation of this infinitely detailed world. This adequacy is clearly a matter of degree and information. In the context of MCDM, the progression of the decision maker may see a rapid 'assimilation' of information about the problem context and, more importantly, about underlying preferences. While the consequence of this evolution may be a reduction of imprecision, the uncertainty about the nature and evolution of knowledge is a notion which engulfs both risk and imprecision, and it is this uncertainty which we term vagueness.

The distinction between imprecision and vagueness is admittedly a tenuous one. Imprecision is intended as a lower level uncertainty governing feelings of doubt and inadequacy in the current state of the decision process. Vagueness, on the other hand, is an almost unperceived governing force that determines the limits of the entire decision process. Although uncertainties relating to the structuring of the MCDM problem are not of primary importance, many may be captured by the notion of vagueness, including all the modelling uncertainties in French's taxonomy. It is a catch-all which provides a header of 'given the totality of information at the time' to the analysis. Belief in a situation in Zimmerman's taxonomy, vagueness in Klir and Folger's taxonomy, and uncertainty about the evolution of future beliefs in French's taxonomy are all encompassed by vagueness.

2.4 Uncertainty in MCDM: A Morphological Box

A morphological box is a crosstabular approach which can be employed to tease apart the important aspects of the problem at hand. Two fundamental 'axes' comprising the current problem are identified: one relating to the nature of uncertainty present in the decision, and the other relating to multiple criteria aspects of the same decision. Section 2.3 characterised three categories of uncertainty which are considered to be most relevant to MCDM. Section 2.1 presented three broad methodological MCDM schools in a deterministic context. Although neither classification can claim to be ei-

ther exhaustive or disjunctive, they do provide a basis for the following morphological framework for the consideration of the treatment of uncertainty in MCDM methods.

	Risk	Imprecision	Vagueness
Value Function Methods			
Outranking Methods			
Metric Methods			

We have been able to find references in the literature to all but two of the boxes, namely those representing vagueness in the value function methods and vagueness in the metric methods. All the other boxes have been considered to a greater or lesser extent.

We may consider risk to be an ‘external’ uncertainty in that it essentially relates to the environment or decision context locating the decision problem i.e. is independent of the DM. Imprecision and vagueness, on the other hand, may be grouped together as ‘internal’ uncertainty which is specific to a particular DM. It is obvious that the three uncertainties are not mutually exclusive, so that more than one type may be present in a single problem context. In fact it has been argued here that vagueness, since it represents uncertainty about the current state of knowledge, will always be present in a decision problem, and the only issue is whether we wish to model it explicitly or not. A high-level treatment of uncertainty in MCDM therefore might consist of all three types of uncertainty considered simultaneously. This thesis does not pretend to have such grand ambitions: it presents only a brief discussion of the key issues in currently popular models of imprecision and vagueness, focusing primarily on the treatment of external uncertainty.

The focus on risk is due to two main considerations. Firstly, the time and resource constraints on the thesis preclude the consideration of all three uncertainties. There is also a very real sense of building towards an integrated model, of ‘learning to walk before running’. Risk then, although not occurring in all decision problems, is the most visible uncertainty when it does occur. While the internal uncertainties of imprecision and vagueness may be ameliorated through discussion and restructuring, risk is unlikely to be as flexible. These two points in particular motivate the consideration of the single uncertainty, risk, represented by the first column in the morphological box.

	Risk	Imprecision	Vagueness
Value Function Methods	Ch. 3.1	-	-
Outranking Methods	Ch. 3.2	-	-
Metric Methods	Ch. 3.3	-	-

2.5 A Note on Fuzzy Sets

The fuzzy set theory introduced in 1965 by Zadeh [110] is currently the most complete and well-known approach for the general modelling of imprecision. The fundamental notion in fuzzy sets is that imprecision manifests itself as an arbitrariness in establishing precise boundaries for a set of interest. This allows the membership of an element x to a set A to be considered a matter of degree by allowing a membership function $\mu_A(x)$ to take on any value between 0 and 1, which are respectively equivalent to the classical concept of set exclusion and set inclusion. By defining appropriate sets of interest, fuzzy set theory is immediately and broadly applicable. Several interpretations are possible depending on the definition of the set of interest. For example, we could define a set 'fast cars'. A car with a top speed of 180 km/h would belong to this set to a greater extent than a car with a top speed of 170 km/h, so that we may assign membership values of 1 and 0.8 respectively. Otherwise, in evaluating car A by its top speed, we may define a set of interest 'top speed of car A '. It is then possible to reflect uncertainty about the exact top speed of the car by assigning membership values of 1 to 180 km/h and 0.8 to 175 km/h if it is believed that 180 km/h is more representative of the car's top speed than 175 km/h. In this sense, fuzzy MCDM corresponds more to a sensitivity analysis embedded into the decision model than anything else.

It is our purpose here to discuss the relevance of fuzzy sets to MCDM, rather than to present the various methodologies in any detail. Generally, the aim of a fuzzy MCDM analysis is to provide fuzzy global evaluations that will inform the DM about the range of global scores that may be obtained given the imprecise information. Naturally, some evaluations will be considered more representative of an alternative than others, so that we may attach a membership function to the fuzzy global evaluations of each alternative. In contrast to conventional MCDM providing crisp evaluations, ranking the fuzzy global evaluations is non-trivial due to the possibility that the evaluations overlap each other considerably, so that several different ranking methods exist [16]. An overview of the huge amount of research done into fuzzy mathematics is presented by Kandel [44]. Chen and Hwang [16] provide a detailed survey of available fuzzy value function and outranking methods. In the value function methods, different solution methods obtain the membership function of the global evaluations either following Baas and Kwakernaak's sequential evaluation of α -cuts [4] or Bonissone's use of approximate arithmetic operations on fuzzy numbers [8]. Greater acceptance of fuzziness has been achieved in the outranking methods, largely through the influence of the well-known ELECTRE III [75] and PROMETHEE [10] methods which use thresholds to allow the concordance and discordance indices to take on values in the interval [0, 1]. A more complete treatment of imprecision has recently been presented by Goumas and Lygerou [31]. The original fuzzy GP method is sometimes credited to Zimmerman [113], although the first explicit formulations were proposed by Narasimhan [64] and Hannan [35]. Generally, the fuzzification of the constraints jus-

tifies them being treated equivalently to goals, so that the fuzzy decision is computed on the basis of a maximin aggregation over the confluence of goals and constraints.

The intuitive appeal of including the natural imprecision of human perception into the decision making process means that fuzzy set theory has received a great deal of attention, perhaps disproportionately so. Nevertheless, there are still doubts as to some fundamental issues at the heart of the use of fuzzy sets, particularly for the purposes of decision aid, to which we now turn our attention.

The Cognitive Relevance of Fuzzy Sets

Perhaps the most important of these doubts relates to the suitability of fuzzy sets as normative models of decision making. The implicit assumption is that the imprecision of qualitative human judgment is well modelled by fuzzy mathematics. However, a well-founded axiomatic system on the basis of which we may judge this assumption is still lacking [27]. Even the simple operations of intersection and union have no well-defined conceptual parallels, so that several definitions have been proposed [114]. A related problem is the lack of behavioural justification for the many ranking methods available, which is particularly important since different methods may give different results. The maximin operation used in the metric methods can be considered overly pessimistic and undesirable in many contexts [71]. French [27] discusses some of the difficulties in fuzzy measurement theory.

The use of fuzzy MCDM techniques relies heavily on membership functions, yet their elicitation is not clear, and typically seem to have been chosen for convenience on the basis of minimum and maximum values, a value with maximum membership, and linear interpolation between the points to form a triangular fuzzy number. Again, whether this structure is consistent with decision maker preference is not apparent, as is the impact of different forms of the membership function on results. Some indication might be taken from results that similar triangular approximations of probability distributions (as in PERT, for example) performed very poorly [46].

Restrictions on the Fuzzy Elements of MCDM Models

Generally, different MCDM methods consider only certain elements of the decision problem to be fuzzy. We may consider the 'external' aspects of the decision problem (evaluations, probabilities), and the 'internal' aspects of the DM (marginal value functions, thresholds, goals, criterion weights) as potentially fuzzy elements of an MCDM model. Value function methods have treated both the evaluations z_{ij} and criterion weights w_j as fuzzy. Having recognised the fuzziness of trade-off information, it is then ignored by demanding the precise specification of the marginal value functions. The opposite is true in the outranking methods, which model the imprecision of DM preferences via the pseudocriteria, but consider both the evaluations z_{ij}

and the weights w_j to be crisp.

The apparent contradiction is that having recognised the imprecision of human judgment we must pick and choose certain sources of imprecision while ignoring others. But which forms of imprecision are most keenly felt by DM's, and which forms have the most influence on results? Arguably the area where DM's are likely to feel the imprecision in their preferences most acutely is in the definition of the marginal value functions, particularly considering the importance of correct modelling and the large impact of inaccuracies on results, reported by Stewart [85]. However, Goumas and Lygerou [31] rejected the use of fuzzy thresholds in the PROMETHEE generalised criteria on the basis that the resulting preference index may spread out to include almost the entire interval $[0, 1]$. Similar results might be expected for value functions. The requirement of additivity for criterion weights dictates that the fuzziness in one criterion cannot be incorporated without consideration of all criteria simultaneously. A similar argument holds for the scenario probabilities, so that the modelling of these quantities is considerably more difficult. It is certainly true that the multiplication operation means that using fuzzy numbers in several areas of a problem will widen the bounds of the results. The question, however, is whether the decision maker feels that the results are a true reflection of the imprecision or merely spurious effects of overemphasising the imprecision in too many areas.

Facilitation of Learning

From the perspective of practical decision aid, the normative applicability of fuzzy sets is largely a function of the insight and opportunity for reflection provided to the DM. The DM unfamiliar with numerical evaluations and general elicitations may well derive some comfort in the knowledge that the imprecision experienced is actually being accommodated, but whether the final fuzzy evaluations allow for the impacts of different imprecisions to be teased apart is doubtful. There is the risk, when including a fuzzy analysis as part of an MCDM model, that the imprecision compounds itself in such a way as to obscure the various aspects of a decision problem. Since the goal of MCDM is to bring the DM to a better understanding of the problem through learning and focusing on key aspects, it may be more suitable to resolve imprecision at the outset via the elicitation of crisp inputs (with the knowledge that these will be investigated as part of the later sensitivity analysis) rather than attempting to model the imprecision and carry it throughout the process. In many cases, initial imprecision may be at least partly reduced by simplification, restructuring and reflection, analogous to the construction of a value tree decomposing high level criteria that could be interpreted as fuzzy into lower-level, crisp ones.

The difficulty of selecting one crisp number is in any case not completely nullified by the consideration of the fuzzy set, since those same difficulties may arise in the

choice of the parameters and form of the fuzzy number. Fuzzy set theory is a precise mathematical framework which, although allowing for imprecision to impact on results, still demands rigour and exactness from the DM. We may of course define further fuzzy parameters on the original fuzzy number, quickly leading to a problem of infinite regress. At what stage does the imprecision cease to represent a material concern of the DM and become just a spurious property of the model?

Despite these problems and challenges, fuzzy set theory addresses a source of uncertainty that is not conducive to stochastic modelling. These imprecisions might be investigated as part of a sensitivity analysis, but if they are of such magnitude or importance that the decision maker feels paralysed to continue before their inclusion, then fuzzy set theory is at present the only established means of doing so. There seems to be a common misconception that fuzzy sets are to be used as alternatives to probability distributions [108, 14]. The notion of imprecision relates to the frailty of human judgement and imprecise expression which is present regardless of future states of the world. It is difficult to interpret a probability distribution around a value as anything other than uncertainty about the future states of the world, essentially external to the preferences of the decision maker, as much as that is possible in the MCDM context. Rather, the use of fuzzy sets should be viewed independently of the use of probability distributions as modelling the imprecision that is very much inherent and personal to the DM.

2.6 A Note on Rough Sets

The model of vagueness based upon the rough set theory proposed in 1982 by Pawlak [67] is an attempt to capture the essence of uncertainty regarding the state of our available knowledge. Rough set theory is founded on the philosophy that there is a certain amount of information associated with every object in the universe. The assumption is then that objects are considered and described only by the information that is available about them, which to all practical purposes is incomplete. Two objects that are different in the idealised world may therefore appear to be the same in the reduced context. The reduction of classificatory ability is said to occur via the *granulation* of knowledge, leading to objects characterised by the same information being viewed as *indiscernible*. That such granulation occurs is beyond question, and is in part the reason behind the development of the outranking school, where incomparability of alternatives is allowed based on a current and inferior state of knowledge. Even within the utility school, the reliance of the DM on the current state of knowledge is implied via the emphasis on learning processes and facilitating DM understanding. Furthermore the basic tenet of the rough set philosophy, that our representation of our world i.e. the decision problem, is limited by the availability of knowledge, is appealing.

The development of the interaction between rough sets and MCDM is very much in its infancy. Until recently, work was based exclusively on the original framework laid down by Pawlak [68], which assumed the selection of a classification problematique (see, for example, [70, 69]). In 1999, Greco et al. proposed extensions to that framework based on the pairwise comparison of alternatives using graded dominance relations [32], which allowed the ranking problems on which MCDM is founded to be modelled. This recent movement represents a significant advancement. Even more so than fuzzy sets, rough set theory offers the opportunity to model a type of uncertainty that has long been acknowledged in the cognitive sciences, but more importantly one which has no other representation in MCDM. It is our purpose here to provide a discussion of the issues surrounding the integration of rough sets and MCDM, without presenting the theory itself. The original work of Pawlak [67] remains the standard reference on the subject, while Greco et al. [33] provide the most complete review of the current movement. Exemplary applications may be found in [81, 22, 82]. We focus here on three key issues: the role of rough sets in MCDM, the extension of rough sets to MCDM methods, and the interaction between vagueness and other types of uncertainty.

The Role of Rough Sets in MCDM

The most conspicuous aspect of applying rough set theory to MCDM is the role played by the exemplary decisions. Most applications of rough set theory emphasise a prescriptive approach by using a training set to induce a set of decision rules to apply to a wider set of alternatives [82, 22]. This in fact places rough sets within the paradigm of preference disaggregation [39], which seeks to infer a preference model based on existing decisions. This poses few methodological concerns in AI fields such as expert systems and machine learning, where the starting point is that the DM is an acknowledged expert. In the context of MCDM, however, the unfamiliar strategic problems to which MCDM is often applied means that there may be no experts. There appears to have been no discussion as to the desirability or even feasibility of the use of exemplary global decisions. Researchers frequently point to the fact that people are more confident exercising their judgements than explaining them as a psychological justification for using exemplary decisions [32]. From a descriptive point of view, a set of exemplary decisions may well provide the reasons for those decisions and an adequate prediction of future choice. From a normative standpoint, however, there is little to suggest that such confidence is justified. Furthermore, there has been no research as to the effects of flawed exemplary decisions.

Consequently the role of rough sets in MCDM needs to be re-evaluated. Decision aid in general performs the dual purpose of explaining the context of the decision problem to the DM, and prescribing a sensible course of action given that particular decision context. The great advantage of rough sets is that they provide a unique depiction of the former, using very natural concepts like the classificatory ability of attributes.

Most importantly, these computations do not require any further preference information if they are performed directly on the decision table of evaluations. A rough set approach is in this regard ideally suited to a pre-processing stage, where the set of attributes identified as most significant by the rough set analysis may be modified based on preference information elicited from the DM before moving into a deeper analysis. If the focus of interest is shifted away from a preoccupation with prescription, the rough set theory offers detailed and mature opportunities for a greater understanding of the decision context at a time of the process which is often neglected.

Extensions of Rough Sets to MCDM Methodologies

The relational basis of rough set theory has resulted in its development in MCDM being purely along the lines of the relational outranking methods. The rough set approach extends the information obtained from the intuitive global evaluations of the DM to the full decision problem. In this sense, the outranking model appears embedded in the rough set approach. In doing so, the contextual information described previously is expressed, with additional information contained in the decision rules. Although these decision rules provide a straightforward interpretation of the reasoning behind the rank ordering which is both valuable and easy to understand [70], their derivation and the consequential prescription does require substantially more DM involvement. Most importantly, all information except that relating to the decision rules is obtained without recourse to the outranking methodology. Whether the information contained in the decision rules is considered sufficiently desirable to employ the full rough outranking approach is likely to be dependent both on the context of the problem and whether the outranking method is considered by the analyst as appropriate.

At present, the role of rough set theory in value function models is limited to information gathering in a pre-processing phase. The challenge here is to fuse together the functional aspects of the value function methods with the relational aspects of rough set theory. The link between discrete metric methods and rough sets is likely to be similar in that a decision table of deviations may be considered. However, metric methods may derive greater benefit from a rough set analysis on the basis of shared interests in pre-processing and similarly low levels of DM involvement. A general issue that has not received treatment is the distinction between the importance of attributes from the perspectives of rough set theory and MCDM. The importance weights in MCDM are pieces of preferential information elicited from the DM that, while depending on the range of values for a particular attribute, are also dependent on what the DM considers to be valuable. As a result they may not be derived directly from the evaluations. The importance of attributes in the rough set theory is purely a function of the classificatory ability of each attribute, requiring no input from the decision maker. Currently, there is no treatment of preferential importance in the derivation of decision rules.

Integration with Other Uncertainties

The interaction between the various forms of uncertainty provides another rich and important area for study. There has been some theoretical research in the area combining rough and fuzzy sets [24], most notably providing fuzzy extensions of similarity and dominance relations. Although the motivation for relationships such as similarity and dominance to be considered a matter of degree is clear, the introduction of fuzzy sets does not alleviate the problems associated with rough set applications to MCDM outlined here. If anything it may provide unnecessary sophistication to an already complex process. The consideration of stochastic outcomes together with rough sets has received far less attention, perhaps suffering from the relative scarcity of stochastic outranking methods. In the investigative pre-processing environment advocated, there is no reason why, where the uncertain outcomes have a discrete representation, scenarios should not be treated in the same way as attributes. Given the evaluations of n alternatives on m criteria for p scenarios, the rough set investigation may be setup using either $p \times n \times m$ alternative-criteria information tables or $m \times n \times p$ alternative-scenario information tables. The combined rough set analysis may provide information on crucial or redundant scenarios, on which scenarios are associated with a particular alternative or attribute, and dependencies between scenarios on each attribute. In a prescriptive environment, Zaras [111] provides a modified pairwise comparison table based on the concept of stochastic dominance, on which the conventional rough set techniques may be applied.

There are therefore numerous directions for future research. However, for rough set theory to make significant inroads into general MCDM practice, there needs to be a general shift of emphasis into softer procedural issues such as ease of use, interactivity and transparency. This represents a fundamental change from the current focus on more sophisticated mathematics and algorithms on the part of rough set practitioners based predominantly in AI.

Chapter 3

External Uncertainty: The Treatment of Risk

The uncertainty arising from stochastic outcomes, which we have termed risk, has received more attention than other forms of uncertainty. It is widely accepted, not only in MCDM but in the context of broader mathematical modelling, that uncertainty arising from such physical randomness should be modelled using the tools of probability theory [28]. The starting point here is taken as the conclusion of the problem structuring phases, already well into the decision aiding process. If the resultant attribute values z_{ij} are dependent on the outcome of some future random event(s), they may be viewed as random variables with associated probability density functions $f_{ij}(z_{ij})$ and cumulative probability distributions $F_{ij}(z_{ij})$. However, the methodological differences between MCDM methods mean that the incorporation of probabilities into the techniques often proceeds in very different ways. In this section, an examination is made of the use of probability distributions in each of the three MCDM areas. We make one important distinction: between methods making use of the full probability distribution and methods using only a subset of the full range of outcomes, which we term scenario-based approaches.

In section 1.6 we defined a scenario as an internally consistent future state of the world that is also in some sense a characterisation of the full multivariate probability distribution governing the performance evaluations, the aim being to simplify what is likely to be an extremely complex functional form. There are two fundamentally different approaches to the construction of scenarios in a decision process under uncertainty. The first is similar to what Keeney has termed (and criticised as) 'alternative-focused thinking' [47], where the scenarios are viewed as joint realisations of the stochastic z_{ij} , implying that the alternatives must be considered first. Subjective probabilities are often assigned to such realisations. The set of scenarios is usually a small subset of the complete set of possible future states, and so the probabilities are usually interpreted as relative likelihoods, although this is often not made explicit. An advantage of such an approach is that it does allow the DM to investigate the range of possible

outcomes that may follow the choice of a given alternative. However, Wright and Goodwin [106] argue that the scenarios generated will tend to suffer from well-known heuristics such as anchoring and adjustment [94], and inertia [37]. As a result the scenarios tend to reinforce the currently held perceptions of what is likely to occur, so that excessive confidence is placed in a narrow range of possible futures.

The second approach has been applied predominantly in strategic planning, and was developed primarily through pioneering work done at Royal Dutch/Shell [101, 100] in response to failures of probability-based forecasting techniques. Scenarios are constructed in a top-down approach that focuses on the underlying forces causing uncertainty in order to identify and differentiate predetermined and uncertain elements of the future via cause-and-effect relationships. The approach aims at identifying an alternative that is robust to the range of scenarios that will bound the uncertain future, but it is the focus on these cause-and-effect relationships which gives attention to the sources of uncertainty without assigning probabilities to them. The scenarios are developed prior to (or at least independently of) the construction of alternatives, so that the process emphasises the articulation of values and objectives rather than charting the possible courses of alternatives. The placement of this construction approach is therefore in the framework of Keeney's 'value-focused thinking' [47], albeit applied to the treatment of uncertain outcomes.

The two approaches differ substantially in their philosophical approaches to uncertainty: the first attempts to manage the uncertainty via its representation as probabilities, while the second may be placed in the framework of the so-called soft OR approaches by emphasising the incorporation of uncertainty as a structure of the problem and an opportunity to better understand the contextual environment in which the problem has been formed, rather than as a solution method. However, these differences will generally manifest themselves in the creation of scenarios (see, for example, [11]), while from an MCDM perspective a major challenge is how to incorporate the *already formed* scenarios into an analysis, and particularly to aid in the evaluation phase. Thus while acknowledging the importance of appropriate scenario construction, it is possible to move directly into areas where the two schools of thought share common procedural elements.

An important objective of scenario-based MCDM is evaluating performance under each scenario, which is considered to be a deterministic set of evaluations. The output of a scenario-based MCDM approach must include separate evaluations for performances under each scenario if it is to contribute to conclusions of robustness. This motivates the treatment of scenarios as part of the objectives hierarchy, which naturally brings about questions of where in the hierarchy to place them. Consideration of the objectives hierarchy as a framework reducing composite objectives into

lower-level objectives clarifies the issue to some extent. The demand that aggregation across scenarios be delayed until the end if it is to be carried out at all means that it is natural to include the scenarios as the second level of the hierarchy. This has the effect of creating a ‘super-MCDM’ problem consisting of p generally closely related problems. Including the scenarios at lower levels will necessarily require aggregation across scenarios in order to progress to further evaluations.

There are two basic modes of evaluation which are common to the scenario-based implementation of all three of the MCDM methodologies considered in this thesis. For each criterion j , alternative-scenario combinations can be formed by considering the performance of alternative i on scenario k . These alternative-scenario combinations can then effectively be treated as higher-level alternatives, so that the evaluation of the set of n alternatives on criterion j occurs over all scenarios simultaneously. Goodwin and Wright [30] term this ‘the evaluation of *alternative-scenario combinations over each criterion*’. Otherwise, for each criterion j , the alternatives may be evaluated separately within each scenario, which is termed here the evaluation of *alternatives over each criterion-scenario combination*. We therefore refer to the domain of the scenario-based decision problem as being either the criterion space (in the former case of evaluating alternative-scenario combinations) or criterion-scenario space (in the latter case of evaluating alternatives). These modes of evaluation are discussed in more detail in the following sections.

3.1 Value Function Methods

3.1.1 Multiple Attribute Utility Theory

Multiple attribute utility theory (MAUT) is the oldest and most well-established of the frameworks provided to cope with stochastic outcomes in MCDM, and despite numerous severe and long-standing objections, at least to its empirical validity, it remains a standard textbook treatment of risk. In fact, a surface scan of risk treatment in MCDM might well lead one to believe that the subject begins and ends with MAUT, although fortunately this is not the case. In this section, the aim is to present MAUT from a broadly structural point of view in order to gain insight into the way that uncertainty is firstly incorporated into the model, and then handled in the progression towards a final decision. The reader is referred to Keeney and Raiffa [48] for a more comprehensive coverage of the subject.

The multiple attribute value theory (MAVT) discussed in 2.1.1 can be viewed as a special case of MAUT by considering that the z_{ij} in the former are just degenerate random variables. The outstanding feature of MAUT, both the reason for its popularity and the criticism levelled at it, is that it is firmly founded on the axiomatic base of expected utility theory. These axioms are assumptions made about the be-

haviour of the decision maker in terms of the way that preferences are formed, and are required to justify the existence of a utility function $U(\cdot)$ satisfying the expected utility hypothesis, which is stated below.

Expected Utility Hypothesis

$$a \succ b \iff E[U(a)] > E[U(b)]$$

Let $X = \{x_1, \dots, x_r\}$ be a set of possible prizes, while the lottery offering prizes x_i , $i \in \{1, 2, \dots, r\}$ with associated probability p_i is represented as $\{p_1, x_1; \dots; p_r, x_r\}$. Furthermore, some of the prizes may be entries into further lotteries. The axioms necessary and sufficient for the existence of a utility function are presented as:

Completeness Axiom

$$\forall x_1, x_2 \in X, x_1 \succ x_2 \text{ or } x_1 \prec x_2 \text{ or } x_1 \sim x_2$$

Transitivity Axiom

$$\forall x_1, x_2, x_3 \in X, \text{ if } x_1 \succ x_2 \text{ and } x_2 \succ x_3 \text{ then } x_1 \succ x_3.$$

Consistency Axiom

$$\begin{aligned} \forall x_1, x_2 \in X, x_1 \sim x_2 &\iff (x_1 \succ x_2 \text{ and } x_2 \succ x_1) \\ x_1 \succ x_2 &\iff x_2 \not\succeq x_1. \end{aligned}$$

Continuity Axiom

$$\forall x_i \in X, \exists p_i \in [0, 1] \text{ such that } x_i \sim \{p_i, x_1; (1 - p_i), x_r\}$$

Independence Axiom

$$\forall x_1, x_2, x_3 \in X, x_1 \succ x_2 \iff \{p, x_1; (1 - p), x_3\} \succ \{p, x_2; (1 - p), x_3\}$$

The completeness axiom requires that the decision maker be able to express some opinion when confronted with a choice between two lotteries: either a strict preference or indifference. It demands a certain level of decisiveness of the decision maker, which has been argued, particularly in support of the outranking methods, to be too strong to be a valid axiom [77]. Although it may be true that, at least early in the decision process, preferences are incomplete enough to justify a conclusion of ‘incomparability’, supporters of MAUT require only that in theory a decision maker will be able to balance the available information at a certain stage in the decision process to arrive at a decision of weak preference [27]. Transitivity of preferences is a more natural expectation, although transitivity of indifference is often systematically violated in practice i.e. it is descriptively invalid [83]. The violation hinges on the

decision maker's lack of discriminatory power, so that from a normative point of view the axiom may be viewed as rational and relatively uncontroversial. The consistency axiom places a rational bound on the range of preferences that can be held, and the continuity axiom is a technical requirement that, although descriptively dubious, can be justified in a normative sense by the same arguments as for transitivity. The crucial utility-theoretic axiom is that of independence, which demands that preferences between lotteries exclude common outcomes from consideration. The independence axiom has an abstract normative appeal, but in practical examples has been systematically and consciously violated by a wide variety of decision makers, giving rise to several well-known paradoxes [1, 25].

From a descriptive standpoint, the violations merely confirm the descriptive failings of expected utility theory, and are only threatening from a normative perspective if they can be justified in a 'rational' sense. A common justification includes the so-called framing effects of reference levels against which decision makers compare alternatives [12], so that the way in which an attribute value is viewed depends on the context in which it is presented. This explanation is often supplemented by considering how uncertainty is incorporated into the final decision i.e. the role played by the probabilities [107]. The EUT axioms demand that probabilities are viewed purely as weights to be attached to the prospective rewards of the lottery. In practice, the probabilities may be viewed as more complex measures of risk, so that the relative weightings implied by two probabilities do not necessarily capture the DM's true feelings. Furthermore, the normative appeal to ignore these perceptions of the probabilities as irrational does not seem entirely convincing, and runs the risk of alienating decision makers unable to reconcile themselves to the axiomatic requirements of EUT. The alternate treatment of the probabilities lie at the heart of the relatively recent developments in non-expected utility theory, which will be discussed in more detail later. However, it must be added that the non-expected utility-theoretical models come with problems of their own, having generally traded off broad applicability and ease of use for increased realism. In addition, fatal violation of EUT is likely to occur only in relatively few cases. Firstly, the majority of decision makers are at least in tentative agreement with the axiom of independence [27], and secondly, some apparently non-EU behaviour can be classified as decision making with error [13] rather than axiomatic violation of EUT.

When extending the ideas of the previous section into a multiple criteria framework, an aggregation of marginal utility functions for individual criteria is required. In the case of MAVT, the acceptance of the preferential independence axiom was required to justify additive aggregation in the construction of the global value function. However, the only demand on the global value function was that it be order preserving. Where outcomes are uncertain, the additional requirement that the expected utility hypothesis be satisfied demands a different form of independence

between criteria. It is the precise nature of these independence assumptions that motivate the use of a specific functional form for the global utility function. In this section the various assumptions of independence are traced through to their corresponding global utility functions. We make use of the notation $u_j(a)$ to refer to the utility of a specific alternative a on criterion j .

Utility Independence A criteria space is defined as $z_1 \times z_2 \times \dots \times z_m$ and partitioned into (\mathbf{X}, \mathbf{Y}) . Then if preferences between lotteries over (\mathbf{X}, \mathbf{Y}) for fixed \mathbf{Y} depend only on the marginal probability distributions of \mathbf{X} and not on the fixed level of \mathbf{Y} then the attributes in \mathbf{X} are utility independent of the attributes in \mathbf{Y} . If the reverse is also true i.e. \mathbf{Y} is utility independent of \mathbf{X} , then mutual utility independence holds between the attributes in \mathbf{X} and \mathbf{Y} .

If each attribute is utility independent of the remaining attributes, but full mutual independence does not hold, then the utility function has the multi-linear form:

$$U(a) = \sum_{j=1}^m \kappa_j u_j(a) + \sum_{i=1}^m \sum_{i < j < m} \kappa_{ij} u_i(a) u_j(a) + \dots + \kappa_{12\dots m} u_1(a) u_2(a) \dots u_m(a) \quad (3.1)$$

If the attributes are mutually utility independent for all partitions (\mathbf{X}, \mathbf{Y}) then (3.1) becomes the special case of the multiplicative utility function:

$$1 + \kappa U(a) = \prod_{j=1}^m [1 + \kappa w_j u_j(a)] \quad (3.2)$$

$$\text{where } 1 + \kappa = \prod_{j=1}^m [1 + \kappa w_j] \quad (3.3)$$

Additive Independence Additive independence is said to hold if preferences between lotteries depend only on the marginal probability distributions of the z_j , such that for example, $\forall \mathbf{x}, \mathbf{x}' \in \mathbf{X}$ and $\forall \mathbf{y}, \mathbf{y}' \in \mathbf{Y}$,

$$\{0.5, (\mathbf{x}, \mathbf{y}); 0.5, (\mathbf{x}', \mathbf{y}')\} \sim \{0.5, (\mathbf{x}, \mathbf{y}'); 0.5, (\mathbf{x}', \mathbf{y})\}$$

If attributes z_1, z_2, \dots, z_m are additively independent, the global utility function can be expressed as the special case of (3.2) when $\kappa = 0$ in an additive form:

$$U(a) = \sum_{j=1}^m w_j u_j(a) \quad (3.4)$$

Just as independence assumptions have implications for the functional form of the utility function, various attitudes towards risk also exert an influence. The two most

important issues are whether the DM always prefers the expected value of a lottery to the lottery itself i.e. is risk averse for all outcomes, or vice versa i.e. is risk prone for all outcomes, and the strength of the aversion or proneness to risk. It can be shown [48] that a DM who is risk averse always has a concave utility function, while a risk prone DM's utility function is always convex. The modelling of risk attitudes into the decision process is therefore appealing both as an aid to consistent evaluation and as an opportunity for the decision maker to better understand the nature of his or her preferences and the implications for choice.

The axiomatic base of MAUT has allowed for a comprehensive study of model properties, so that the theoretical basis of MAUT is well established, yet there is a perceived gap between theoretical development i.e. traditionally academically orientated research, and widespread practical usage [109]. A brief example of a hypothetical decision process expresses the difficulties that may give rise to this gap. For the multiplicative model (3.2), the final choice is that alternative maximising expected utility i.e.

$$\max_i \mathbb{E} \left[\prod_{j=1}^m (1 + \kappa w_j u_{ij}) \right] \quad (3.5)$$

The calculation of expected utility requires knowledge of all cross-product moments up to order m . In most practical cases the evaluation of marginal probability distributions is a time-consuming and difficult task; it is difficult to conceive of widespread use of any method that places even more arduous demands on its users. A natural question is then to ask to what extent simplifications to the utility model affect the integrity of the results. A persistent and fundamental contradiction exists [85] in the general advancement of methods designed to incorporate the violations of rationality axioms, and the apparently inconsequential results of simplifications that in effect 'induce' violations. As an approximation of (3.5) one may reverse the order of product and expectation:

$$\mathbb{E} \left[\prod_{j=1}^m (1 + \kappa w_j u_{ij}) \right] \approx \prod_{j=1}^m \left(1 + \kappa w_j \mathbb{E}[u_{ij}] \right) \quad (3.6)$$

This simplified representation requires only marginal probability distributions, and effectively makes the assumption that the criteria are stochastically independent. A further simplification is to reduce the model to an additive form:

$$\prod_{j=1}^m \left(1 + \kappa w_j \mathbb{E}[u_{ij}] \right) \approx \mathbb{E} \left[\sum_{j=1}^m w_j u_{ij} \right] \quad (3.7)$$

Note that the second simplification (3.7) excludes three aspects of the multiplicative representation from consideration: the unit constant, the multiplicative factor κ and

all the higher order terms. The first two exclusions imply that the u_{ij} of the multiplicative model and the u_{ij} of the additive simplification are not necessarily the same. However, they are related through a positive affine transformation, and are therefore strategically equivalent i.e. they represent the same set of preferences. Other effects of both simplifications were analysed by Stewart [85], who found the effects to be minimal in terms of incorrect representation of decision maker preferences, particularly when contrasted with those errors introduced by incorrect assessment of the marginal utility functions. Where preference reversals did occur, alternatives were generally so similar that any sensitivity analysis would have detected the near indifference. This suggests that a simplified representation coupled with a detailed sensitivity analysis will in nearly all cases provide the same level of decision aid as more complex models.

3.1.2 Scenario-based Approaches

The role of the probabilities was vital in bridging the gap between deterministic MAVT and stochastic MAUT. The aggregation of the uncertain outcomes into expectations implied further requirements in order to satisfy the expected utility hypothesis. The analogous choice in a scenario-based value function approach is therefore a vital but seldom stated one: do we want to aggregate over scenarios using probabilities to find expected utilities? Advocates of scenario planning [96, 101, 106] would say not, emphasising the consideration of the totality of information in search of an alternative which performs well in all eventualities. In such a case the desired output is p rank orders, implying that a value theory framework is sufficient. However, many practical cases do use scenario probabilities to form expectations, and Pomerol [71] has argued to the effect that the maximin criterion implied by scenario planners is overly pessimistic.

In any case it does not seem unreasonable to attempt to aggregate over scenarios in some way. In a small sized problem the formulation giving p complete rank orders may be enough to arrive at a decision, but for any moderate to large size problem the interpretation of the $n \times p$ matrix of evaluations may benefit from some form of summary. The nature of the summary is largely dependent on the needs of the decision maker: it may take the form of an expected utility, but also might include other weightings, graphical aid, variance measures or deviations from some reference point if these can be specified. The latter aggregations all imply a value theoretical treatment. In cases where expected utilities are computed, we must operate in the utility theory framework defined by the usual Von Neumann-Morgenstern axioms. This will have implications for the construction of the utility functions as well as the aggregation. However, for ease of presentation in this section we refer to both as value functions, except where confusion may arise.

The construction of the global value function itself involves some important practi-

cal issues raised by Stewart [87]. The assumption of preferential independence was required to ensure the validity of the additive representation. Within the context of conventional (not scenario-based) MAVT it is advisable to restructure a problem when preferential independence is found not to hold between two or more criteria. The question as to whether there are *a priori* reasons to expect the scenarios to cause preferential independence violations cannot be answered in a general sense, although the actual assessment of violations may prove more difficult when scenarios are incorporated.

A potential problem associated with the use of scenarios may arise out of causal (rather than preferential) dependencies where the relative tradeoffs between two criteria are viewed differently under strongly diverse scenarios. We conjecture that the additional qualitative information discovered during the construction of scenarios might potentially change the tradeoff preferences of the decision maker substantially. Where alternatives have been evaluated over each criterion-scenario combination, any potential problems associated with capturing changing preferences over scenarios would be avoided, since each scenario is considered separately. Preferences, for the criterion in question, would be represented by p separate marginal value functions, not necessarily all different. This offers a potential solution where such structural problems have occurred in practice. Another related suggestion might be to use an alternative elicitation technique in the rating of the alternative-scenario combinations – for example the direct rating method.

The construction of different marginal value functions v_{jk} for each criterion-scenario combination has not been directly proposed in the literature, although there have been some allowances for changing preferences over scenarios in other areas [51]. It is helpful to attempt to reconcile the ideas of MAUT with the proposed construction. Typically, evaluations of values are performed by asking lottery-type questions for a criterion j in order to generate a marginal value function v_j . Then, a swing weight w_j is defined by considering the relative importance of, for example, the extreme points of the function relative to those of the other criteria. What does this assume? We may certainly write out a marginal value function in each scenario; although different scenarios might operate in different regions of the value function, the global form of the marginal value function may be assumed to remain the same. Now, if the value to the DM of a change in the underlying attribute from 0 (or the global minimum) to x in all scenarios is not the same for all x , there is no justification for the single value function over criterion j i.e. there exists a one-to-many relationship between the underlying attributes and the values. This implies that the evaluation of alternative-scenario combinations over each criterion (as in MAUT) is valid only when

$$w_{j1} = w_{j2} = \dots = w_{jp}, \forall j \quad (3.8)$$

We conjecture that where an alternative-focused construction of scenarios occurs, the marginal value functions are elicited with respect to an implicit status quo or reference scenario. It may well occur that as a result the functions elicited from the DM do not capture the full range of preferences. The extent to which this is true and affects the progression of the analysis is difficult to answer. The detailed cause-and-effect structure of the scenarios allows for the construction and exploration of additional qualitative information. The question then is to what extent the additional information influences the preferences of the DM, or at least the awareness of those preferences. A follow-up question is whether the influence is sufficient to warrant the increased complexity of allowing preferences to change over scenarios. Decision makers may in general struggle to consider attributes within the context of a given scenario, particularly when that scenario may be extreme. Certainly, MAUT represents an already complex approach for facilitated decision aid. It is important not to provide total accommodation of aspects that in many cases might have little or no effect on the success of the analysis.

In both cases the choice of domain (criterion-scenario or just criterion) for the value function implies a natural consequence for the assessment of weights. If alternative-scenario combinations have been rated over each criterion, then a 0–100 swing in value will incorporate all p scenarios, so that assessments of weights by questions like ‘Is a 0–100 increase in criterion 1 preferable to a 0–100 increase in criterion 2?’ naturally elicit a $m \times 1$ vector of joint weightings, which can be used to compute aggregate evaluations for alternative a in each of the p scenarios by

$$V_k(a) = \sum_{j=1}^m w_j v_{jk}(a) \quad (3.9)$$

Note that the weighting vector w_j is constant over all p scenarios, due to the evaluation of the alternative-scenario combinations over each criterion incorporating the relative importance of each scenario in the definition of the partial value functions, and consequently also in the joint weightings.

On the other hand, if alternatives have been rated over each criterion-scenario combination, the weighting issues become considerably more complex. A 0–100 swing in value will incorporate only one scenario, so that two lines of questioning can be followed: the first attempts to simultaneously capture importance information on both criteria and scenarios i.e. to elicit the joint weights directly, while the second elicits separately the importance information for criteria and scenarios before aggregating them into a joint weighting. It is helpful at this stage to return to the value tree formulation discussed on page 31. There we suggested that the scenarios be placed in the second-highest level of the value tree as parents to p structurally similar ‘within scenario’ value trees. For ease of presentation we assume that the within-scenario value trees are identical, which might often but not always be the case, so that there are mp

lower-level criteria. We may then consider the distinction between the elicitation of *cumulative* and *relative* weights discussed in [7]. Cumulative weights are assessed over all bottom-level criteria simultaneously, and are traditionally normalised to sum to 1. Relative weights, on the other hand, are assessed within families of criteria i.e. those criteria sharing the same parent, and are normalised to sum to 1 within that family. Depending on the levels of the value tree and the consequential choice of parent, different types of relative weights can be elicited. For our purposes, we consider only the choice of scenarios as bases for the relative weighting of bottom-level criteria. From an algebraic point of view, the use of relative and cumulative weights is equivalent. The cumulative weight of a bottom-level criterion is simply the product of its relative weight in comparison to its siblings and the relative weight of its parent, in this case the scenario to which it belongs.

In the scenario-based decision problem, the direct joint weighting is a rather straightforward extension of the elicitation of cumulative weights that considers each criterion-scenario combination as a bottom-level criterion, so that we may ask ‘In which criterion-scenario combination is a 0–100 increase most desirable?’. Although technically this may appear simple, practically it is likely to be far less so. The elicitation procedure demands that DM’s weigh up different criteria and different scenarios simultaneously, which might be particularly difficult if the two are not independent i.e. if some criteria become relatively more important under different scenarios. Furthermore the elicitation of all mp criterion-scenario weights is likely to prove tedious for even moderate-sized problems.

In response to these difficulties we may turn to the second elicitation approach, in which information about the relative importance of criteria and scenarios are elicited separately. In this process we initially consider each of the p scenarios separately, eliciting relative criterion weights within each scenario using questions such as ‘Is a 0–100 increase in criterion 1 preferable to a 0–100 increase in criterion 2 under scenario 1?’. The proposed line of questioning addresses the relative importance of each criterion, but no scenario information is elicited. The result is a set of p $m \times 1$ vectors of relative criterion weights ψ_{jk} , one vector for each scenario, indicating the weight of criterion j under the conditions of scenario k . Naturally it is possible to consider further breaking down the within-scenario weight elicitation by considering intermediate level criteria to be parents of the bottom-level criteria and children of the scenarios. However, we do not lose any generality by assuming that this is not the case, and rather that the relative criterion weights sum to 1 within each scenario.

In order to arrive at a cumulative weighting it is still necessary to obtain an estimate of the relative importance of each scenario. The elicitation and interpretation of weights for higher-level criteria is not straightforward. Although in comparing two

higher-level criteria the DM should technically be considering the relative effect of a simultaneous 0–100 swing on *all* sub-criteria of the higher-level criteria, this often gives way to the use of more abstract notions of importance in the interests of practical ease. Such practice leads easily to confusion about the meaning of importance and subsequent inconsistency. Instead the relative scenario weights may be quite easily obtained by extracting one criteria from each scenario and comparing them in the usual swing weighting sense. We would suggest the choice of the same criterion j in each scenario in order to focus the DM’s attention on the fact that it is scenario weights that are being elicited, although there are no algebraic reasons why different criteria cannot be used. The elicitation process can and should be repeated with other criteria as a consistency check. The choice of the criteria used should ultimately be based on the ease with which the DM is able to think about the available tradeoffs, and is at the discretion of the analyst. In any case, the result is a set of p relative scenario weights ϕ_k . The joint or cumulative weighting w_{jk} can be found by multiplying the relative scenario weights ϕ_k by the relative criterion weights ψ_{jk} .

$$V_k(a) = \phi_k \sum_{j=1}^m \psi_{jk} v_{jk}(a) \quad (3.10)$$

The weighting vector is now different over the p scenarios, but the results of the computations remain aggregate evaluations for alternative a in each of the p scenarios. The relative importance of the scenarios is now taken into account through the calculation of the scenario weights ϕ_k rather than through the partial value functions. An important point is that if all that is desired is a rank ordering in each of the p scenarios, we need not even consider the scenario weights. The tradeoff between simplicity of the value function and concise elicitation of importance weights is clear: using the criterion domain requires complex partial value function construction but relatively simple weight elicitation, while using the criterion-scenario domain requires more demanding weight elicitation, but simpler constructions of the partial value functions. Both may be used simultaneously as a consistency check, although in most practical circumstances this will be too time-consuming to be a feasible option.

As a final point on the issue of weight elicitation, we mention the possibility of a second approach for eliciting the relative weights. Instead of first eliciting the importance of criterion j given scenario k , we may invert the process to consider the weight of scenario k given criterion j , using questions of the form ‘Is a 0–100 increase in criterion 1 more preferable under scenario A or B ?’. This relative scenario weighting would be followed by the elicitation of relative criterion weights by extracting one scenario from each parent criterion, analogously to the process already outlined. Although either weighting approach is technically valid, we restrict further attention to the latter approach on the basis that it appears more compatible with scenario thinking by obtaining information *within* each scenario.

As mentioned previously, the final form of the aggregation is an open question depending largely on the goals of the DM. It is important to draw the distinction between the notions of robustness implied by scenario planning and the compensatory aggregation implied by more traditional MCDM approaches. The search for a robust alternative suggests some sort of maximin aggregation, although to avoid overly pessimistic behaviour it may be beneficial to consider ‘approximately robust’ alternatives [76], that is alternatives performing satisfactorily in *almost* all scenarios. The idea of good performance in all scenarios is perhaps unrealistic in the context of general MCDM problems, but the negligibility of poor performance is dependent on the DM and the problem faced. The extent to which this should and does occur are important questions. Lehman [55] has provided some answers to the former, defining a negligible scenario as one in which any changes in the corresponding outcomes are insufficient to change the DM’s preferences for a particular alternative.

Within the traditional forms of aggregation, it is possible to obtain a global value score for each alternative by summation over the p scenarios

$$V(a) = \sum_{k=1}^p V_k(a) \quad (3.11)$$

Note that no additional weighting is required regardless of whether (3.9) or (3.10) is used. Both approaches have already incorporated the relative importance of each scenario in the construction of the V_k . If probabilities are used,

$$V(a) = \sum_{k=1}^p \text{Pr}[k] V_k(a) \quad (3.12)$$

We have made use of Stewart’s result to the extent that an additive approximation may be made with little detriment to the results [85]. Note that this relates only to the form of the value function, and not to the approximation abilities of the scenarios themselves. A troubling aspect is the lack of interpretation for the scenario probabilities in any natural sense, although often relative likelihood is used as a surrogate measure [2, 57]. This may not be such a problem if one is willing to accept the narrowing of scope that the approximation in (3.12) implies. If relative likelihoods are used, the optimal alternative is merely that one which maximises utility over the p scenarios. It must then be accepted that if the choice of scenarios does not adequately bound the set of all future states i.e. none of the scenarios occur, then the choice may have been suboptimal. This emphasises the importance of careful scenario generation.

3.1.3 Non-Expected Utility Theory

The so-called non-expected utility methods were predominantly developed as economic models of individual choice under uncertainty, with some assistance from the

psychological modelling of potential heuristics used by decision makers. The emergence of the non-expected utility models in the late 1970's was an attempt to reconcile the tractability of optimising a single utility function with the axiomatic violations, particularly of independence, which occurred in practice. As such, the focus is far more on a descriptive theory of decision making taking place in a single dimensional framework, which takes this section out of the scope of this thesis to some extent. However, in light of the earlier discussion on the possible advantages of incorporating these violations into a decision model, it is instructive to see how these accommodations have been made. This section is limited to a discussion of two of the more accepted and useful non-expected utility models: prospect theory is strongly focused on descriptive accuracy, while rank-dependent expected utility is, as its name suggests, proposed as an extension of EUT. Starmer [83] gives details of around ten more non-expected utility theories.

There are two fundamental concepts in the non-expected utility models. One is the notion of a reference point against which outcomes should be evaluated in order to allow for differential treatment of perceived gains and losses [12]. Then, subjective attitudes towards probabilities are often responsible for the psychological weight attached to an event, called the decision weight, differing from the probability of that event. Wu and Gonzalez [107] offer evidence of this, as well as the two possible explanations of subadditive probability judgments and subjective weighting for the nonlinear relationship. A nonlinear decision weighting function may cause problems eliciting utilities using standard procedures such as the certainty equivalent method, since those questions posed to the decision maker may no longer have the same interpretation for analyst and decision maker. In general, the distortion of probabilities overweights low probability events while underweighting high probability events, resulting in an 'inverted s-shaped' decision weighting function that is concave below an inflection point and convex above.

One of the earliest and most popular models was the prospect theory proposed by Kahneman and Tversky [41] in the field of applied psychology as a procedural representation of decision making as a two-stage process. The first stage comprises an editing phase in which alternatives are subjected to a number of decision heuristics in order to simplify the evaluation and selection tasks of the second phase. In this second phase, a choice is made using evaluations of alternatives via a utility function $U(a) = \sum_i \pi(p_i)u(x_i)$, where the $\pi(\cdot)$ allows for alternate weighting of probabilities and the outcomes x_i are interpreted as gains or losses relative to a reference point. The shape of the utility function is prespecified to be concave for gains, convex for losses and steeper in the domain of losses. These properties are interpreted as implications of diminishing marginal utility and loss aversion. The inverted s-shaped weighting function is advocated based on the assumption that the end points of the probability scale i.e. 0 and 1, represent natural reference points so that diminishing

marginal sensitivity implies steepness around these points. Prospect theory suffers from a number of faults, particularly from a normative point of view. An application of prospect theory may result in the selection of stochastically dominated alternatives due to the direct transformation of probabilities into decision weights causing non-monotonicity in $U(\cdot)$. This is a particularly damaging property which is raised by many as a fatal objection, even in a descriptive context [83]. The selection of dominated alternatives is addressed by including the deletion of all detected dominated alternatives in the editing phase, which then gives the editing phase the appearance of a safety net, albeit at the beginning of the process. In addition, from any practical perspective the detection of dominance will be a difficult or impossible task. Nevertheless, desirable features such as nonlinear decision weights and reference levels were included in later construction of more conventional non-expected utility models.

Rank-dependent expected utility theory (RDEUT) was proposed by Quiggin [73] and has been described as ‘the most natural and useful modification of the classical expected utility formula’ [58]. The key feature of the RDEU model is its use of cumulative probabilities in the calculation of the decision weights to ensure monotonicity of the utility function. This has the very appealing property that the weight assigned to an outcome depends not only on its true probability, but on its ranking relative to other outcomes. If outcomes are ranked from worst to best, x_1, \dots, x_r then weights are given by

$$\begin{cases} w_i = \pi(p_i + \dots + p_r) - \pi(p_{i+1} + \dots + p_r) & \text{for } i = 1, \dots, r-1 \\ w_i = \pi(p_i) & \text{for } i = r \end{cases} \quad (3.13)$$

Rank-dependent expected utility theory comes closest to addressing the true concerns about EUT by transforming the probabilities into measures that may more fundamentally represent decision maker sentiment when making choices under uncertainty. The inverted s-shaped weighting function $\pi(\cdot)$, which overweights low-ranked outcomes relative to higher-ranked outcomes, has the interpretation of pessimism, and shares a close association with risk aversion in the utility function [83]. From an axiomatic point of view, RDEUT relaxes the independence axiom of EUT by demanding that preferences between lotteries be unaffected by substitutions of common outcomes that leave the rank order of outcomes in both lotteries unaffected, a condition known as co-monotonic independence.

As mentioned in the introduction to this subsection, the non-expected utility theory methods have been developed in the context of maximising a single utility function i.e. for a single criterion decision problem. While the extensions and modifications proposed by the non-expected utility theorists allow for a more detailed modelling of preferences, they are often considerably more complex than the model suggested by EUT. In the context of *multicriteria* decision modelling, each criterion would need to be extended to incorporate non-expected utility characteristics, so that the complex-

ity of the model would increase rapidly in the number of criteria considered. It seems plausible to suggest that this perceived complexity may have prevented non-expected utility theory from making a greater contribution to MCDM.

3.2 Outranking Methods

The development of the outranking methods has to a large extent proceeded independently of any treatment of risk. Although some attempts have been made to incorporate stochastic evaluations into an outranking process, there appears not to have been any discussion as to what the desirable aspects of such a treatment are. As a result, the proposed models give the impression of improvisations rather than structured attempts to bridge the gap between deterministic and stochastic models. This is a direct consequence of the well-known lack of axiomatic development in the outranking methods [97].

3.2.1 Pairwise Reduction of Probability Distributions

The evaluation of each alternative i on criterion j results in a probability distribution f_{ij} defined over the random variable z_{ij} representing the relevant attribute value. The methods described here reduce the problem to the pairwise comparison of these distributions, so that in the spirit of outranking one may say whether the probability distribution f_{aj} is at least as preferable as the distribution f_{bj} . As mentioned, the methods differ considerably with respect to the manner of comparison.

In 1977 Jacquet-Lagrèze [38] proposed the first treatment of distributional evaluations in the context of outranking. The method consists of a two-stage process. In the first stage, the intersection of the two probability mass functions is removed and interpreted as evidence in support of indifference. In the second stage, the remaining probability mass on each event is examined via the cumulative distributions and portioned out as evidence either that alternative a is preferred to b , or that alternative b is preferred to a . In a setting where probabilities are defined over a discrete event space $E = \{e_1, \dots, e_h, \dots, e_p\}$, an algorithm is provided consisting of setting up a $p \times p$ crosstabulation of the two probability mass functions f_{aj} and f_{bj} . Row totals obtained by summation over columns thus represent the probability of obtaining precisely the event e_h for alternative b i.e. $f_{bj}(e_h)$, while column totals represent the probability of obtaining the event e_h for alternative a , i.e. $f_{aj}(e_h)$. The diagonal elements provide evidence that a is indifferent to b , lower triangular elements provide evidence supporting the preference of a over b , and upper triangular elements provide evidence supporting the preference of b over a . It is then possible to aggregate each measure over all criteria by using a weighted summation of the individual preference measures, analogous to the deterministic aggregation of ELECTRE I. Although the method is algorithmically simple, conceptually it is far less so. The elements of the

cross-tabulation appear to have no concrete interpretation outside of their membership to a specific type of evidence; certainly their probabilistic meaning is unclear. Then, the incorporation of preference thresholds will considerably complicate the method conceptually and computationally. Again, lack of firm physical interpretations make it difficult to see how this interesting method might be reconciled with more modern outranking approaches.

A second suite of models [21, 61, 62] compares distributions by constructing a matrix \mathbf{P}^j whose entries P_{ab}^j denote the probability that alternative a is superior to alternative b on criterion j i.e. $\Pr[z_j(a) \geq z_j(b)]$. The models differ with respect to the subsequent exploitation of the probabilities.

Dendrou et al. [21] simply aggregate the P_{ab}^j over criteria to arrive at a global index for each pairwise comparison

$$P_{ab} = \sum_{j=1}^m w_j P_{ab}^j \quad (3.14)$$

where w_j is interpreted by the authors as the probability that criterion j is decisive in the assertion that alternative a is superior to b . It is not clear what, if any, advantages this operationally difficult interpretation has over the traditional voting one. The probabilities can be represented in a matrix \mathbf{P} . The global pairwise comparisons are further aggregated according to the joint probability

$$\gamma_a = \prod_{b \neq a} P_{ab} \quad (3.15)$$

A rank order is based on descending values of γ_i . Note that the evaluations in (3.15) imply that if an alternative a is dominated by any other alternative, then $\gamma_a = 0$. Finally, the model presented here can be linked to traditional ELECTRE models by defining an ‘agreement matrix’ \mathbf{P} and a ‘disagreement matrix’ $\mathbf{1-P}$, corresponding to notions of concordance and discordance respectively.

Martel et al. [61, 62] extend the previous approach to incorporate more sophisticated measures of outranking. A confidence index $c_j(a, b)$ indicating the support for the statement that a outranks b is defined over P_{ab}^j , taking into account various indifference and preference thresholds. Then a local doubt index $d_j(a, b)$ corresponding to discordance is constructed in a two-stage process. In the first stage, a disagreement index is defined with the following desirable properties

- Increases with the expected value of the excess of $z_j(b)$ over $z_j(a)$.
- Increases with the weight w_j attached to criterion j .

- Decreases with the dispersion of the distributional evaluations for $z_j(a)$ and $z_j(b)$.

In the second stage, the disagreement index is augmented with indifference and preference thresholds to form the local doubt index. Note that the distributional aspect of the problem is fully absorbed into the problem by this stage through the definition of the P_{ab}^j and the disagreement index. The aggregation and exploitation of the confidence and doubt indices may proceed as for the deterministic ELECTRE III model. Two issues should be noted. Firstly, the thresholds model the ambiguity or fuzziness of the DM's preferences; this model is therefore one of the few to integrate more than one type of uncertainty into an MCDM method. Then, thresholds must be defined either in terms of probabilities P_{ab}^j or the even more abstract disagreement index. It is an interesting behavioural question whether the DM is able to work with these concepts to confidently specify the thresholds.

The third and most recent group of techniques for the pairwise comparison of probability distributions was proposed by Zaras and Martel [112], and uses concepts of stochastic dominance. For a single criterion, stochastic dominance is a collection of pairwise distributional comparisons that can be used to evaluate the expression that alternative a is at least as good as alternative b . The stochastic dominance relations are defined as

$$F_{aj} >_1 F_{bj} \iff H_1(x) = F_{aj}(x) - F_{bj}(x) \leq 0, \forall x \in [0, \infty) \quad (3.16)$$

$$F_{aj} >_2 F_{bj} \iff H_2(x) = \int_0^x H_1(y) dy \leq 0, \forall x \in [0, \infty) \quad (3.17)$$

$$F_{aj} >_3 F_{bj} \iff H_3(x) = \int_0^x H_2(y) dy \leq 0, \forall x \in [0, \infty) \quad (3.18)$$

where the relation $>_i$ is referred to as first-, second-, and third-degree stochastic dominance for $i = (1, 2, 3)$ respectively. By assuming different classes of utility functions, it is possible to express preferences consistent with the expected utility hypothesis, in terms of only stochastic dominance relations.

$$a >_1 b \iff E[U(a)] > E[U(b)], \forall U_1 \quad (3.19)$$

$$a >_2 b \iff E[U(a)] > E[U(b)], \forall U_2 \quad (3.20)$$

$$a >_3 b \iff E[U(a)] > E[U(b)], \forall U_3 \quad (3.21)$$

where U_1 is the class of all increasing utility functions, U_2 is the class of all utility functions that are concave and belong to U_1 , and U_3 is the class of all utility functions that are decreasingly risk averse and belong to U_2 . Similar conditions have been provided for convex utility functions [111].

When moving into a multicriteria framework, Huang et al. [36] have shown that a

necessary condition for multiattribute stochastic dominance is stochastic dominance on each individual criterion. Given the conflicting nature of multicriteria problems, this is hardly an operationally useful result. The authors therefore propose the use of a weighted aggregation

$$C(a, b) = \sum_{j=1}^m w_j c_j(a, b) \quad (3.22)$$

where $c_j(a, b) = 1$ if a stochastically dominates b on criterion j , and is otherwise zero. This results in a concordance index as for ELECTRE I. In particular, this allows for the specification of only a subset of attributes via the choice of appropriate thresholds. More recently, Zaras [111] has proposed a rough set approach to finding a suitable set of criteria to model multivariate stochastic dominance.

Martel et al. [60] and Azondékon et al. [3] have considered stochastic dominance as a more general measure of preference, by defining a local preference index

$$c_j(a, b) = \begin{cases} \varrho_j(a, b)\zeta_j(a, b)\theta_j(a, b) & \text{if } \exists i : a >_i b \\ 0 & \text{otherwise} \end{cases} \quad (3.23)$$

where $\varrho_j(a, b)$, $\zeta_j(a, b)$, and $\theta_j(a, b)$ are functions scaled between 0 and 1 that decrease as dominance conditions weaken from $>_1$ to $>_3$. Thus $\varrho_j(a, b)$ attains its maximum only when f_{aj} lies completely above f_{bj} , $\zeta_j(a, b)$ whenever $a >_1 b$, and $\theta_j(a, b)$ whenever $a >_2 b$. The previous method can be considered the special case of $\varrho_j(a, b)\zeta_j(a, b)\theta_j(a, b) = 1$. The graded preference is in accordance with the increased difficulty of perceiving preferences between distributions for which weaker dominance conditions hold. Local indices can be aggregated into a global outranking index using (3.22), following which the matrix of indices may be exploited using an ELECTRE III or PROMETHEE procedure. A general problem relating to the practical use of the stochastic dominance methods is the rather inaccessible concepts relating to the definitions of dominance, particularly of the second- and third-order. There is a danger that stochastic dominance is viewed as a black box, so that the explanation and justification of results and the iterative nature of the decision process is likely to be adversely affected.

3.2.2 Construction of a Distributive Outranking Relation

In response to the early synthesis of the distributional evaluations performed in the previous models, d'Avignon and Vincke [19] proposed a model which, rather than summarising the stochastic evaluations as P_{ab}^j , uses them to proceed to a distributive outranking degree indicating the probability of attaining various degrees of outranking.

A preference index $I_j(a, b)$ denoting the preference for alternative a over alternative

b on criterion j is defined for all possible combinations $(z_j(a), z_j(b))$ of the stochastic evaluations and scaled between 0 and w_j , where w_j is the weight attached to criterion j once all weights have been normalised to sum to one. A degree of preference $H_j(a, b)$ indicating the preference for alternative a over b on criterion j can then be associated with the probabilities that $I_j(a, b)$ takes on various values. Provided that the evaluations of the alternatives can be considered independent, the probability corresponding to each $H_j(a, b)$ is given by

$$\Pr[H_j(a, b) = h] = \sum_{\{(e_r, e_s): I_j(a, b)=h\}} f_{aj}(e_r) f_{bj}(e_s) \quad (3.24)$$

where $f_{aj}(e_r)$ is the probability of obtaining event e_r for alternative a on criterion j , and $f_{bj}(e_s)$ is the probability of obtaining event e_s for alternative b on criterion j . The key feature of the method is the construction of this random variable $H_j(a, b)$. The aggregation of the preference degree into a distributive outranking relation is made more difficult by the distributional aspect of the problem, but proceeds in the spirit of the ELECTRE III and PROMETHEE approaches. Essentially the preference degree $H_j(a, b)$ may be aggregated into a pairwise outranking degree by simple addition over criteria,

$$S(a, b) = \sum_{j=1}^m H_j(a, b) \quad (3.25)$$

before distributive measures of average ‘strengths’ $S(a)$ and ‘weaknesses’ $W(a)$ are defined as

$$S(a) = \frac{1}{n-1} \sum_{b \neq a} S(a, b) \quad (3.26)$$

$$W(a) = \frac{1}{n-1} \sum_{a \neq b} S(b, a) \quad (3.27)$$

Again, the exploitation of these measures is far more difficult due to their distributional nature. An initial suggestion [19] is to use the median of the distributions to arrive at a rank order with respect to each of strengths and weaknesses. More advanced exploitation procedures have been proposed in [19]. Although the distributive method is impressive from theoretical and methodological points of view, it is a complicated algorithm composed of parts for which no firm interpretations seem available. Its general practical applicability remains untested.

3.2.3 Scenario-based Approaches

As for the value function models, the use of scenarios in outranking applications of the scenario approach may proceed in two methodologically different ways, based on whether or not we attempt the construction of an outranking relation and subsequent

rank order in each scenario, to be tentatively aggregated in the final analysis. There is little in the literature to suggest any work on either of these approaches. As mentioned before, models treating the problem of uncertainty in the outranking methods tend to be isolated from one another, so that there is little unified effort.

Nevertheless, the methods presented in sections 3.2.1 and 3.2.2 all involve direct use of user-specified probability distributions. From a methodological point of view, it makes no difference whether the full or reduced probability distribution is used, so that these methods may be applied without modification. There are, however, practical implications of such scenario-based reductions which raise important questions. Both the number of scenarios and the way in which scenarios are selected are likely to strongly influence the results. This is also true for the value theory models, but is especially clear in the context of methods making pairwise comparisons of probability distributions i.e. the methods of Jacquet-Lagrèze [38], Martel et al. [61], and Zaras et al. [112]. Whether the direct comparison of probability distributions make these methods more sensitive to scenario selection is a question that could be investigated reasonably easily via simulation. It would appear crucial to maintain the same spread as the full probability distribution, perhaps in the form of some extreme quartiles. Such a strategy would coincide with the traditionally popular approach of selecting the best- and worst-cases before (qualitatively) interpolating between them to select the remaining $p - 2$ scenarios. The work of Keefer and Bodily [46] in approximating probability distributions is likely to be particularly relevant to the outranking methods in this context.

If we wish to take a different approach to risk by building up a deterministic model in each scenario, then it becomes necessary to view the scenarios as part of the objectives hierarchy. As discussed earlier, the scenarios fit naturally into the second level of the hierarchy, giving an MCDM super-problem consisting of p structurally similar problems. The issues of interest are the definitions of the concordance and discordance measures, the assessment of weights, and the manner of aggregation. The method of evaluating alternatives over each criterion-scenario combination will proceed in a similar way to deterministic outranking methods. Concordance and discordance measures on each criterion can be computed separately for each scenario, although the precise nature of the computations will depend on which outranking method is employed. The ELECTRE I measures are given by

$$C(a, b) = \frac{\sum_{(j,k) \in \Delta(a,b)} w_{jk}}{\sum_{j=1}^m w_{jk}} \quad (3.28)$$

$$D(a, b) = \begin{cases} 1 & \text{if } \exists(j, k) : z_{jk}(a) - z_{jk}(b) > t_{jk}^v, \\ 0 & \text{otherwise.} \end{cases} \quad (3.29)$$

where $\Delta(a, b)$ is the set of all criterion-scenario combinations for which a is indifferent or preferred to b , and t_{jk}^v are veto thresholds for criterion j and scenario k . The ELECTRE III measures are defined as

$$c_{jk}(a, b) = \begin{cases} 1 & \text{if } z_{jk}(a) + t_{jk}^i \geq z_{jk}(b), \\ \frac{z_{jk}(b) - (z_{jk}(a) + t_{jk}^p)}{t_{jk}^i - t_{jk}^p} & \text{if } z_{jk}(a) + t_{jk}^i \leq z_{jk}(b) \leq z_{jk}(a) + t_{jk}^p, \\ 0 & \text{if } z_{jk}(a) + t_{jk}^p \leq z_{jk}(b). \end{cases} \quad (3.30)$$

$$d_{jk}(a, b) = \begin{cases} 0 & \text{if } z_{jk}(b) \leq z_{jk}(a) + t_{jk}^p, \\ \frac{z_{jk}(b) - (z_{jk}(a) + t_{jk}^p)}{t_{jk}^v - t_{jk}^p} & \text{if } z_{jk}(a) + t_{jk}^p \geq z_{jk}(b) \geq z_{jk}(a) + t_{jk}^v, \\ 1 & \text{if } z_{jk}(b) \geq z_{jk}(a) + t_{jk}^v. \end{cases} \quad (3.31)$$

where t_{jk}^i, t_{jk}^p and t_{jk}^v are the indifference, preference and veto thresholds respectively, so that $t_{jk}^i < t_{jk}^p < t_{jk}^v$. The notion of discordance may prove more problematic in a scenario context. Specifically, discordance means that if performance is poor enough on a criterion, *even in one scenario*, then that outranking relation is vetoed. This implies a strict condition, so that care should be taken that the veto thresholds are consistent with such an interpretation. In particular, it appears that performances in a single scenario would need to be extreme in order to incur a veto.

The question of changing preference structures over scenarios is raised briefly once again. In a manner similar to the specification of the marginal value functions, it is theoretically possible to specify different threshold values in each scenario. Since all thresholds may be functions of z_j , the scenario-based modifications are matters of preference rather than the attribute domain of a particular scenario. It is again an interesting behavioural question whether certain circumstances can be imposed which might lead the DM to consider tightening or loosening thresholds depending on which scenario is considered. As we are considering each scenario independently of the others, weights are permitted to differ across scenarios to reflect different importance judgements by the DM. Similarly to the elicitation of weights discussed on page 38 for the scenario-based value function method, weights may either be assessed as the joint importance w_{jk} of criterion j under scenario k i.e. direct elicitation of the cumulative weights, or as the product of the relative criteria weights ψ_{jk} and relative scenario weights ϕ_k . The different interpretation of the weights attached to each outranking criterion-scenario imply the need for different assessment procedures. In particular, the relative scenario weights might proceed as before by extracting one criterion from each scenario, or could assess the (essentially abstract) importance of each scenario directly. The technical incorporation of scenario-dependent weights is secondary to considerations of ease of use and interpretability. As before, it is not clear whether decision makers are able to isolate their preferences conditional on a

particular scenario. Furthermore, where this is possible, it is not clear how demanding this task is, and to what extent it is a desirable aspect of a decision aid (in terms of insight gained and effect on results).

The outranking method at this stage consists of a set of m concordance measures $c_{jk}(a, b)$ and m discordance measures $d_{jk}(a, b)$ for each scenario, each considering the statement ‘alternative a outranks alternative b on criterion j in scenario k ’. The degree of aggregation depends on the demands of the DM. If all that is required is an evaluation within each scenario, then the construction of an outranking relation in each of the p scenarios can proceed as for the deterministic outranking methods. That is, we may define an aggregate concordance measure

$$C_k(a, b) = \frac{\sum_{j=1}^m w_{jk} c_{jk}(a, b)}{\sum_{j=1}^m w_{jk}} \quad (3.32)$$

that can be compared to the set of j discordance indices in order to build the outranking relation in scenario k . Without trivialising this often difficult step, we can refer to section 2.1.2 and assume that such an aggregation has resulted in a set of p outranking relations. This provides the DM with at least a partial rank ordering in each scenario, which from the perspective of scenario planning is seen as the most important output. We can note that since no aggregation is occurring over scenarios, it is possible to use either the cumulative weights w_{jk} , or the relative criterion weights ψ_{jk} . It is not necessary to explicitly consider the relative scenario weights ϕ_k .

Nevertheless, it might be beneficial to consider further aggregation over scenarios. Such an aggregation would result in an outranking relation considering the proposal that ‘alternative a outranks b ’ in a global sense. The aggregation of concordance measures across scenarios takes the form of the simple summation

$$C(a, b) = \sum_{k=1}^p C_k(a, b) \quad (3.33)$$

provided that the scenario weights have been incorporated into the earlier construction of the $C_k(a, b)$. The treatment of the discordance indices can consider each of the jk criterion-scenario combinations analogously to a criterion in a deterministic outranking context, so that the global discordance measure can be written as

$$D(a, b) = \prod_{j,k} f(d_{jk}(a, b), C(a, b)) \quad (3.34)$$

where

$$f(d_{jk}(a, b), C(a, b)) = \begin{cases} 1 & \text{if } d_{jk}(a, b) < C(a, b), \\ \frac{1 - d_{jk}(a, b)}{1 - C(a, b)} & \text{if } d_{jk}(a, b) > C(a, b) \end{cases} \quad (3.35)$$

The data for the problem therefore resembles exactly the input for a deterministic outranking method. Based on this super-outranking problem a final outranking relation is constructed with respect to the chosen method. For example, the ELECTRE III credibility index is again given by $S(a, b) = C(a, b)D(a, b)$.

An obvious point is that we have not included scenario probabilities at any stage of the analysis. The lack of axiomatic assistance and the behavioural fuzziness around the notions of thresholds and weights makes it difficult to decisively identify where probabilities might be included. The flexibility around the definitions of the weights provides some scope to include notions of relative likelihood as well as relative importance there. This would imply that (3.33) could simply be modified to reflect the inclusion of probabilities. Although the notion of discordance appears to exclude probabilities by much the same justification by which weights are excluded from consideration, it seems likely that the relative likelihood of different scenarios will have an impact on the setting of veto thresholds. In such a relationship, a less likely scenario would require larger differences in performance to induce a veto, all else equal. A more precise characterisation of this relationship, and the abilities of decision makers to think in such a way, are difficult practical questions.

As a final point we return briefly to the problem of evaluating alternative-scenario combinations over criteria. Although this was the approach followed for MAUT, it was shown in section 3.1.2 to imply some quite strict conditions. Within the outranking methods, the direct comparison of alternatives on different scenarios takes the form of questions to the effect that alternative-scenario pair (i, k) outranks (j, l) . This provides additional information reflecting the implied relative importance of each scenario, so that it is unnecessary to consider scenario information in the weight elicitation. However, the pairwise nature of the comparisons in an outranking context may result in unwelcome discordances in this elicitation environment. In particular, it allows for the possibility that two alternatives may veto each other if performances are very different in two scenarios i.e. 'boom' and 'bust' scenarios. These possibilities alone preclude the elicitation from general use.

3.3 Metric Methods

In a deterministic mathematical programming setting it is assumed that all information necessary for representing the decision is known with certainty prior to the solution of the model. In practice it may often be extremely difficult or impossible to provide numerical values for each of the parameters, either due to future uncertainties (e.g. next year's level of rainfall in a reservoir planning problem), unstable or poorly understood relationships (e.g. of tourism on the demand for a product), or a mixture of the two. In problem contexts where this is the case, an approach that then replaces the uncertain parameters with their expected values by making 'reasonable guesses'

may be adequate when combined with careful sensitivity analysis. Nevertheless, this approach is guilty of 'wishing away' the uncertainty inherent in the system, rather than treating the uncertainty directly. Furthermore, where sensitivity analysis indicates that the optimal solution depends strongly on the uncertain values, it makes sense to seek the extra flexibility offered by considering the uncertainty in a more comprehensive way.

When one or more of the parameters of a mathematical program is represented as a random variable, a stochastic program (SP) results. It is assumed that the joint probability distribution function of the random parameters is known, either described via marginal distributions if the random variables are independent or using joint distributions where some dependence exists. In most cases it is assumed that the set of all possible outcomes of the random parameters is discrete.

The formulation of the problem is no longer well-defined in the sense that the interpretation of optimisation cannot be clarified until the stochastic parameters are observed. This leads naturally into potential solution methods for SP problems. It is instructive to consider the area of single objective linear programming (LP), where most of the work in stochastic programming has been done. A preliminary method in that area is to find a decision that will be feasible for all possible realisations of the random parameters. This type of decision is known as a fat decision [42], and will typically prove expensive due to the extreme nature of the constraints placed on the solution. Beyond the fat solution, there are two standard models differing in their approach to incorporating this infeasibility. Recourse models assign a response to the outcome of the random variables, so that it is possible to associate costs with the violations of constraints. The optimisation criterion is then to maximise or minimise the expected objective function value i.e. profit, cost. Chance-constrained models restrict the probability of infeasibility to a certain level, and then optimise the objective function subject to these probabilistic constraints. Further details of these standard treatments, as well as the considerable problem of numerical solution, can be found in any stochastic programming text e.g. [42, 72]. More recently, Liu and Iwamura [56] have proposed a dependent-chance programming model, which optimises the chance of satisfying an objective given that some other uncertain events occur. For example, the authors present an example maximising the probability that costs do not exceed a certain level, subject to the fulfilment of two demand constraints with uncertain constraint coefficients and RHS constraints.

Within the framework of metric methods, treatments of risk have been scarce [6]. A 1998 goal programming review [92] includes only a few lines on the subject, and more recent reviews of goal programming [54] and reference point approaches [104] do not mention stochastic variants at all. By considering the four methods presented

above within the context of metric methods it is possible to understand the reason for the lack of diversity. The key issue is that in moving from an LP framework to a GP framework, the notion of satisficing means that it is no longer necessary to treat objectives and constraints (as they are defined in LP) differently. A fat solution therefore has no real interpretation in GP apart from one satisfying all goals in all circumstances, which in the rare cases of occurrence would most often be an indication of overly unambitious goal setting. Within the context of recourse models, the penalty function used there penalises deviations from constraints, and so fulfils the role of the deviational variables in GP. The recourse formulation is therefore equivalent to the GP formulation minimising the expectation of weighted deviations i.e.

$$\min \left[\mathbb{E} \left[\sum_{j=1}^m [w_j \delta_j]^\alpha \right]^{\frac{1}{\alpha}} \right] \quad (3.36)$$

There appears to be little work done in this regard, although the formulation is intuitively appealing and was suggested in the first exploration of stochastic GP by Contini [17]. The dependent-chance programming model is applied to the GP problem by Liu and Iwamura [56] by placing aspirations on the probabilities of satisfying the various objectives and minimising the deviations i.e. relegating the probability of occurrence to the role of constraint. This is, however, precisely the same as the ideas behind the extension of chance-constrained programming to GP, in which the cost function previously optimised is now included together with the other probabilistic constraints. Aspirations may be specified for some or all of these constraints to change them into goals, leaving only the deviations to be minimised. The chance-constrained goal programming (CCGP) methods have been the almost exclusive focus of interest in the treatment of stochastic outcomes in the metric methods, and it is to these developments that we now turn.

3.3.1 Chance-constrained Goal Programming

A forerunner to the more well-known CCGP models was proposed by Odom et al. [66]. The authors operationalise the idea that the DM wishes to minimise the risk of not achieving the desired goals, using the corresponding probabilities. This results in a maximisation of the joint probability representing the chance of satisfying all goals simultaneously, although a less sensitive aggregation is also proposed based on a weighted summation of the logarithms of the individual probabilities on each objective. The model falls outside of the sphere of goal programming somewhat, since there is no concept of minimising the distance to some ideal. The method here examines the probability that the deviational variable for objective j lies outside of a user-specified interval,

$$\Pr \left[d_j^- \leq \sum_{i=1}^n a_{ij} x_i - b_j \leq d_j^+ \right], \quad (3.37)$$

but does not penalise greater or lesser violations any differently. There is therefore no attempt to minimise the distance to the aspirations set by the DM. This model seems part of a common confusion between linear and goal programming models, linked to the tenuous and often philosophical distinction between constraints and goals. For example, although (3.37) appears as a LP-type objective function, it is possible to think of the formulation as minimising the probabilistic distance to the ideal point representing certain satisfaction of each goal i.e. the point $(0, 0, \dots, 0)$. In fact, the authors do make a later provision for a threshold probability on each objective indicating the minimum acceptable degree of satisfaction, without providing further details. However, by letting these threshold probabilities be goals with corresponding deviational variables rather than constraints, the model becomes identical to the well-known CCGP models developed around that time.

In work between 1978 and 1980 Keown and Taylor [49] proposed a CCGP model to incorporate risk in the RHS constraints by modelling the b_j corresponding to the constraints as random variables. The GP formulation becomes

$$\min \left[\sum_{j=1}^m [w_j \delta_j]^\alpha \right]^{\frac{1}{\alpha}} \quad (3.38)$$

subject to:

$$\Pr \left[\sum_{i=1}^n a_{ij} x_i \geq b_j \right] + \delta_j \geq g_j, \quad j = 1, 2, \dots, m \quad (3.39)$$

$$x_i \geq 0, \quad \forall i \quad (3.40)$$

where g_j is an aspiration for the probability of satisfying objective (constraint) j . The authors provide a procedure which converts each chance constraint into a (linear) deterministic constraint by assuming that the b_j are normally distributed. Only then are the deviational variables δ_j added. The above formulation is therefore slightly contrary to the sequence of the solution algorithm, but provides a better conceptual view of the model. Shortly after the model of Keown and Taylor, De et al. [20] proposed a similar CCGP model that, instead of incorporating stochastic RHS constraints, allowed the constraint coefficients i.e. the a_{ij} , to be represented by random variables. Although the formulation in (3.38) remains the same, the solution procedure is quite different. The authors provide a procedure to convert the stochastic constraints into a deterministic form which is non-linear. This constraint can then be linearised using Naslund's approximation [65] before adding deviational variables to model aspirations.

There are some important features of this model. Firstly, risk is incorporated via the aspirational probabilities g_j , representing the level of reliability desired by the DM

for the achievement of each objective. Increasing risk aversion is therefore associated with higher values of g_j . However, the CCGP models make no attempt to incorporate the magnitudes of any deviations or constraint violations. It seems fundamentally important that both the magnitude and probability of poor performance be included as bases for evaluation – a conjecture supported by several studies into the nature of risk (for example, [79]). Within the LP framework in which the chance-constrained approaches were first developed, probabilistic constraints are applicable because the constraints are supposed to be hard. However, when moving into a GP framework, a straightforward extension of CCLP seems unwise – we must also consider the magnitude of the violations since the b_j are defined only as goals.

Secondly, the CCGP models do not lend themselves to the discrete ‘decision table’ formulation of an MCDM problem. In these contexts, there are no explicit constraints on performances, so that the b_j represent desired levels of performance on each criterion. This places a significantly greater demand on the DM in terms of required preference information. In addition to aspirations for each objective, the DM must also provide the desired probabilities of achieving each aspiration. In the presence of the aspirational probabilities, it is not clear to what extent the attribute aspirations should be considered goals or constraints. There is certainly some relationship between the two aspirations. The attribute aspirations represent a raw level of desired performance, while the probability aspirations represent the desired reliability of achieving these levels of performance. It may be difficult, however, for a DM to separate the two issues of performance and reliability, particularly when the desired performance on an objective may be heightened either by increasing the desired magnitude or probability of achievement.

Even the ability of the DM to think and express preferences in terms of the aspirational probabilities is an open question. Certainly the setting of aspirations based on attribute values is a challenging enough task, and the use of probabilities rather than attribute values will almost certainly complicate the matter. Well-known behavioural research into the heuristics and biases associated with the interpretation and assessment of probabilities (e.g. [30]) would provide some insight into how these complications could manifest themselves. The availability heuristic, used by a DM judging a probability based on the ease with which the associated event is recalled, might lead to overly risk averse behaviour where the consequences of a violation are severe or unpleasant. Anchoring perceptions on the status quo, on the other hand, might lead the DM to underestimate the probabilities of divergent future states, which would result in insufficient risk aversion. Finally, for well-balanced aspirations it is important that the probabilities are not interpreted as lower-limits on performance, as in the chance-constrained LP models, rather than the goals that they are intended as.

3.3.2 Other Stochastic GP Models

The pioneering work of Contini [17] has already been mentioned, but the narrow focus of the work prevents its general applicability. Specifically each objective j is deterministically related to the decision variables with the exception of a random perturbation $\xi_j \sim N(0, \sigma_j^2)$. Donckels [23] constructs a model along the lines of portfolio theory by minimising a weighted sum of the expectation and variance of the deviations. Essentially this model replaces each objective with two sub-objectives; one minimising expected deviations and the other minimising the variance of those deviations. These specific ‘risk’ criteria place the model to a certain extent in the framework of section 3.4, which considers the use of explicit risk attributes. The author admits that ‘a lot of work remains to be done to make the technique operational’. Specifically, no aggregation is provided over different objectives. Such an aggregation seems particularly problematic, as concepts of tradeoffs between two criteria become murkier as a result of the variance terms. It is a practical question whether DM’s are able to consider tradeoffs between expected values and variances on different objectives. In any case, the theoretical formulation can be stated as

$$\min \sum_{j=1}^m w_j^E E[\delta_j] + w_j^V \text{Var}[\delta_j], \quad (3.41)$$

where w_j^E are weights attached to the expected values and w_j^V are weights attached to the variances. As mentioned previously, it may be of more practical benefit to consider the easier problem of minimising only the expected deviations.

Much more recently, Ballesterio [6] has proposed a stochastic GP model also developed along the lines of the mean-variance approaches popular in portfolio theory. In fact, the model can be considered a generalisation of that portfolio model to several attributes. The starting point of the model is the increasing concave utility function defined over the values of the objective $z_j = \sum_{i=1}^n a_{ij}x_i$, where the coefficients a_{ij} are random variables with associated probability distributions. This utility form is then linked with notions of satisficing to formulate the stochastic GP as

$$\min \sum_{j=1}^m w_j R_j(\bar{z}_j) \sigma^2(z_j)$$

subject to:

$$\begin{aligned} \bar{z}_j &\geq g_j, \quad \forall j \\ x_i &\geq 0, \quad \forall i \end{aligned}$$

where w_j is the weight attached to deviations from objective j , \bar{z}_j is the mean value of objective j , $R_j(x) = u_j''(x)/u_j'(x)$ is the local risk aversion coefficient at the point

x , and g_j is the aspiration level for the expected utility $E[u_j(z_j)]$ of objective j .

The formulation takes the form of minimising risk, as represented by the variance, subject to an acceptable average performance on each objective, and can be solved as a quadratic program. Again several points can be made regarding this model. Firstly, the model steps outside the compensatory mode advocated in GP by reinstating hard constraints on the average objective values. Although the general form of the model is an interesting advancement in financial modelling, it is less clear that the behavioural implications of the mean-variance framework are desirable in broader MCDM. In particular, it is doubtful if DM's think in terms of having to achieve a certain goal, and only then consider risk issues.

Secondly, requirements of the DM are in a sense reduced, as the only aspirations required are those representing desired performance on each objective. However, the aspirations are phrased in terms of expected utility, which requires a far greater insight on the part of the DM. Ballestero comments that 'precise knowledge of the DM's utility functions can help the determination of suitable targets', but it is precisely in order to avoid explicit construction of utility functions that metric methods are often employed. In the absence of full specification of the utility functions, the effects of substandard elicitation of the utilities would need to be investigated, although a careful sensitivity analysis of the impact of different aspirations might go some way toward answering this question within the context of a specific problem. Furthermore, the elicitation of the local risk aversion may prove troublesome in some circumstances and difficult for the DM to reconcile with the aims of the analysis, although some elicitation techniques have been described in [6, 5]. Therefore although the model proposed by Ballestero offers an interesting and significant addition with respect to its link to utility functions, it is this very aspect that perhaps disadvantages it in the pre- or early-processing area employing metric methods.

3.3.3 Scenario-based Approaches

Given the scarcity of stochastic GP approaches in general, it is unsurprising that there have been few developments in the area of scenario-based modifications. The chance-constrained GP models presented earlier made use of probability distributions defined over uncertain parameters of the decision problem. It would not affect the formulation of the problem if only some points of the distributions were used, which means that the methods described earlier may be used without further modifications. The issues in such cases relate to the appropriate number of scenarios to use, and the identification of suitable selection methods for the scenarios. The lack of practical usage prevents these questions being considered in any detail here. Within LP applications of stochastic recourse models [45, 102, 18], the construction of the scenarios appears to take place in a fairly *ad hoc* manner.

Recent work by Korhonen [51] provides a stochastic GP model that, rather than being based on chance-constrained ideas, is effectively an extension of recourse models to scenario-based GP. A set of scenarios is defined using scenario planning techniques, after which the implications of the scenarios in terms of various operating parameters are investigated and estimated. Probabilities are assigned to represent the relative likelihood of occurrence for each scenario. Aspirations may be defined separately for individual scenarios, as are weights attached to the aspirations. The model therefore makes allowances for a changing preference structure over the set of scenarios. The model finally minimises deviations from the relevant scenario-dependent aspirations. Kalu [43] provides another non-standard application of GP under risk, although he stresses the impossibility of probability assignment. Regardless, a pre-emptive GP model is solved under various scenarios representing different combinations of operating parameters, with the optimal decision vector given separately for each of the p scenarios.

Although there seems to be no formal theoretical development of such models, the applications provided by Korhonen [51] and Kalu [43] represent a far simpler interpretation of the GP problem with stochastic outcomes than that provided by the CCGP models. In particular, it allows us to retain the same structure found in deterministic GP approaches and merely apply it in each scenario. We therefore devote the remainder of this section to the investigation of such a model, returning temporarily to the discrete case before providing some modifications required for continuous problems.

The basic idea behind the scenario-based goal programming (SBGP) model is to formulate a deterministic GP for each scenario, followed by an aggregation over all p scenarios. The GP formulation for scenario k is given by

$$\min \left[\sum_{j=1}^m [w_{jk} \delta_{jk}]^\alpha \right]^{\frac{1}{\alpha}} \quad (3.42)$$

where w_{jk} is the weight assigned to the deviation δ_{jk} from goal g_{jk} , for each criterion j and scenario k , and α denotes the choice of norm. In the GP framework, there turns out to be no distinction between the evaluation of alternatives over criterion-scenario combinations and the evaluation of alternative-scenario combinations over criteria. For both the value function and outranking methods, the choice of evaluation method depended on the preference structure of the DM, that is whether a value function or outranking relation could be constructed over scenarios or not. If preferences could be expressed with perfect consistency within either evaluation method, then there is no question that the final results would be identical, so that the issues pertain to ease of use and related practical matters. Now within the goal programming framework, the goals that are elicited from the DM represent far weaker preference information than either that used in the construction of the value function or the outranking relation. It is this weakening of the elicited preference information

that results in the inability of the metric methods to incorporate scenario importance information anywhere other than in the weights themselves. The deviations in (3.42) give no indication of the relative importance of each scenario. The manipulation of goals in order to imply some importance information i.e. by increasing the goals on a more important scenario, is nothing more than an *ad hoc* trick, and is likely to confuse what is a cognitively appealing methodology. Furthermore the goals themselves are heavily dependent on the attribute domain, and it may well be difficult to divorce what is related to scale and what relates to importance. In contrast to both the value function and outranking methods, the structure of the problem does not depend on whether the inputs i.e. goals, differ over scenarios. The implication is that the weight assessment stage must include elicitations of both criterion and scenario importance information.

The assessment of the weights may again proceed in either of the two ways outlined for the value function methods on page 38. Either the weights may be elicited simultaneously over all mp criterion-scenario pairs, or relative criterion weights ψ_{jk} and relative scenario weights ϕ_k may be assessed and multiplied together to find the joint weighting w_{jk} . The first assessment method is a cognitively taxing process, while the second method is conceptually easier at the expense of concision. The concision-conception tradeoff is familiar from previous sections, and again both methods may be used as a consistency check.

The application of (3.42) in each scenario results in p rank orders, each scenario being represented by a $n \times 1$ vector of deviations from which a rank order can be trivially obtained. The results comprising the p rank orders and deviation vectors contain important information in themselves. While the rank ordering produced by an aggregation over scenarios should certainly not be excluded, the information contained in the scenario-wise results should be considered as an important output in its own right. This information should be carefully interpreted and scrutinised before proceeding with any further aggregation. Once this has been done, the deviation vector in each of the scenarios may be grouped together to form a $n \times p$ matrix of deviations i.e. precisely the form of the input for a conventional goal program. In this sense the SBGP appears as an extension of the metric method to the case of the super-MCDM problem. A complete metric approach can then be performed on the input table of deviations. Specifically, a metric other than the α used in the p individual problems may theoretically be used to aggregate the super-problem. This presents an interesting opportunity to incorporate the different preference philosophies encompassed by the different metrics.

Formally the full metric method incorporating a second stage comprising the aggre-

gation procedure may be written as

$$\min \left[\sum_{k=1}^p \left[\left[\sum_{j=1}^m [w_{jk} \delta_{jk}]^\alpha \right]^{\frac{1}{\alpha}} \right]^\beta \right]^{\frac{1}{\beta}} \quad (3.43)$$

where α and β are appropriate metrics for the individual and aggregation problems respectively. Any comparisons between the Archimedean and the Tchebycheff aggregation should be mindful that the two aggregations model different behavioural aims. The Tchebycheff approach is associated with notions of robustness and strong performance over all scenarios, while the expected value aggregation is more compensatory in that it searches for alternatives that are on average stronger. Again, which is more appropriate can only be argued taking into account the specific problem context and DM psychology. However, the aim advocated by the scenario planning philosophy, that is the search for a robust alternative that performs satisfactorily under all scenarios, is best furthered by the Tchebycheff super-aggregation i.e. $\beta = \infty$. Such an aggregation will consider the *worst* relative performance across scenarios i.e. in terms of the deviations, to be the indicator of an alternative's merit, however the performance is measured.

The model has thus far been faithful to the scenario planning philosophy of not assigning any form of probabilistic information to the scenarios. However, rather than the absence of relative likelihood, the implication of (3.43) is one of *equal* relative likelihood. This is clear in that the Archimedean super-GP problem i.e. $\beta = 1$, is equivalent to an expected value technique assigning equal likelihood to each scenario. If the DM feels comfortable in making at least some judgements regarding the relative likelihood of each scenario, there seems no good reason why this information should not be used, bearing in mind that the aggregated rank order is not considered to be the only output of the SBGP. This information is easily incorporated into the following form

$$\min \left[E \left[\sum_{j=1}^m [w_{jk} \delta_{jk}]^\alpha \right]^{\frac{1}{\alpha}} \right] \quad (3.44)$$

The expectation aggregation procedure can be considered a more general form of the Archimedean procedure, and superior based on its ability to incorporate likelihood information where available. The Tchebycheff super-aggregation can be similarly adapted, in which case the relative likelihood of scenario k , $\Pr(k)$, acts as a weight for the performance measure obtained from the within-scenario aggregation.

It is not the intention of this section to say that probabilistic information need be *automatically* included in the SBGP model; merely that such a consideration is not out of the question either. In all cases the inclusion or exclusion of relative likelihoods

should be based on (particularly) the aims of the analysis, the preferences of the decision maker and the context of the decision problem. The history of scenario planning is instructive in understanding this. Given that scenarios are constructed to be *plausible* rather than just *possible*, the scale of the long-term strategic planning in large organisations is of such a nature that ruin in any scenario need not incorporate any probabilistic information to be considered further; ruin in any scenario, given that that scenario is plausible, is sufficient warning. However, this may not always or even often be the case. It is evident that the aims and the problem context to which scenario planning was traditionally applied, rather than any general theoretical concerns, led to the exclusion of probabilistic information in scenario planning.

Extensions to Continuous Goal Programming

In the value function and outranking methods, the scenarios manifested themselves primarily in different attribute values between scenarios, which motivated the treatment of each scenario problem separately. In the continuous GP methods, the scenarios manifest themselves primarily through different goals g_{jk} and constraint coefficients a_{ijk} . Since the LP algorithms generate solutions by optimising over a constrained decision space, there is no reason to treat the p decision spaces separately where a robust decision is desired. The notion of robustness requires a prospective 'optimal' solution to perform well on all objectives under any of the scenarios. In the continuous case, consideration of the problem as being composed of p separate subproblems is likely to return p quite different decision vectors. Each of these vectors may perform best in the relevant scenario, but performance on any other scenario is unknown and may in many circumstances be poor.

Within the context of the Tchebycheff framework, an extension of the notion of satisficing to choice under uncertainty requires that the decision maker focuses on attaining an acceptable performance level on the most important scenario-objective combination before considering other scenario-objective combinations. If the problems are considered separately, this dynamic movement between scenarios is not possible, and satisficing can only occur in a local sense. The concepts of satisficing and robustness are thus closely related. In contrast to the value function, outranking, and discrete GP frameworks that generally split the super-MCDM problem into p separate parts, the continuous GP formulation should consider the super-MCDM problem in itself.

The Role of Interactivity

The need for an interactive element is likely to be strongly increased by the inclusion of a scenario-based treatment of uncertainty. Firstly, there is a far greater number of goals that must be specified than for an equivalent deterministic problem, so that tradeoff relationships are necessarily more detailed. Then there remains the fundamental question as to whether decision makers can effectively evaluate their

preferences under the assumptions of different scenarios. There may be considerable uncertainty in the mind of the decision maker as to the implications of a particular scenario for a given objective. A strongly interactive element will allow the decision maker to explore the uncertainty around various scenario-objective relationships in a ‘what if?’ environment, as well as performing the more conventional search of the decision space to assess the tradeoffs available. The avoidance of the difficult assessment of weights may be particularly beneficial in the early stages of the decision process, where the focus is on identifying a set of potentially optimal solutions for closer inspection without spending too much time doing so.

3.4 Explicit Risk Attributes

An alternative treatment of risk is to consider it as a primitive concern, so that the riskiness in an attribute may be characterised based on the reduction of the uncertain outcomes to a single risk measure. This measurement of risk has been a critical issue in portfolio theory in particular, but also in other areas of economics, psychology and decision science.

The motivation for the consideration of the risk attribute is most easily seen within the context of MAUT, although it applies to the other methodologies as well. Risk preferences are built into the utility function and various measures can be derived once the utility function has been elicited (risk premium, risk aversion function). However, the construction of the utility functions is a time consuming and conceptually intensive procedure, so that there is some motivation to search for surrogate attributes that are able to capture risk preferences. The issue is to what extent the consideration of risk measures that essentially ‘fall out’ of the underlying attributes i.e. no judgemental assessments are included, affect the results, in addition to the selection of the appropriate risk measures themselves. The decomposition of the underlying attributes involves both value and risk components. The reason that the value component has not been mentioned until now is that it appears uncontroversial relative to risk, with the use of expected values being widely accepted [79]. This section is therefore focused on the risk component, where the major obstacle remains the confusion arising from multiple conflicting notions about how risk should be defined and modelled [26].

Weber and Bottom [103] provide a review of empirical research on how the attributes of gambles influence risk, from which they conclude the following

- Risk increases with an increase in range, variance or expected loss of a gamble.
- Risk decreases if a constant positive amount is added to all outcomes of a gamble.

- Risk increases if all outcomes are multiplied by a constant positive number greater than one.
- Risk increases if a gamble is repeated many times.

Although these points should be interpreted in a descriptive sense, the lack of a clear conception of risk means that it is no longer clear whether any distinction should be drawn between normative and descriptive processes. It is certainly clear that the notion of risk should incorporate subjective aspects peculiar to specific DM's. Furthermore, different notions of risk might be held by, for example, financial analysts (variation around an expected return) and nuclear safety experts (chance of a catastrophe). This provides the motivation to consider directly the link between measures of risk and preference models. Sarin and Weber [79] summarise previous attempts at using empirically valid definitions of risk and value to construct so-called risk-value models of preference for decisions under uncertainty in the context of lotteries. Therefore for two lotteries X and Y , with the status quo wealth level denoted as W ,

$$X \succeq Y \iff f(V(X), R(X), W) \geq f(V(Y), R(Y), W) \quad (3.45)$$

where V measures the value of a lottery, R measures its risk, and f is a function increasing in V and decreasing in R reflecting the tradeoff between value and risk, and may depend on the level of wealth W . More recently, Jia and Dyer [40] have proposed a measure of risk based on the normalisation of lotteries to have expected values of zero, which allows the risk and value components to be clearly distinguished. The defining characteristic of the so-called standard measure of risk is that it depends on the utility function of the DM. This allows for the preference-based justification of commonly used risk measures as well as, under a condition analogous to utility independence, an alternative representation of the expected utility model that explicitly expresses tradeoffs between value and risk.

There are two results that are particularly relevant to MCDM. Firstly, the use of variance as a measure of risk (at least in a utility theory model) is appropriate only in two situations: where the decision maker has a quadratic utility function $u(x) = x - \beta x^2$, and where the utility function is increasing and concave and the random variables representing lotteries are normally or log-normally distributed. Some numerical support is provided by Stewart [88] by means of a simulation study, where it was found that the variance may be a misleading risk measure when the utility function takes the form of a power function and lotteries are non-normal. However, when an exponential utility function is used, the variance appears to be an adequate risk measure even where lotteries are non-normal. A possible explanation for this may be that the risk aversion function for the exponential utility function closely resembles that of a quadratic utility function (for which lotteries need not be non-normal to validate the variance as a measure of risk) in the time domain specified in the Stewart simulations. The implication of these results is that the popular impulse to associate risk

with variance, inherited from the 1952 portfolio model of Markowitz [59], should be resisted at least until the requirements have been verified.

The second point relates to the use of cumulative probability distributions to provide comparisons in terms of risk, which is intuitively plausible although somewhat different from the notions associated with the variance-based measures of risk. Rothschild and Stiglitz [74] provided a risk measure defined over the cumulative probabilities F and G associated with lotteries X and Y respectively, but which is sufficient only when the lotteries X and Y have the same expected return, so that

$$X \succeq Y \iff \int_0^x (G(t) - F(t))dt \geq 0 \quad \forall x.$$

Stewart also analysed the use of cumulative distributions in the same simulation context [88] and found that ‘there appears a *prima facie* case to be made that the use of two well-chosen cumulative probabilities may be useful’. Again there were mixed results depending on the form of the underlying utility function, and the use of cumulative probabilities was more successful assuming a power utility function relative to an exponential utility function, in contrast to the earlier results for the use of the variance.

Hallerbach and Spronk [34] have suggested a ‘multiple factor approach’ to risk measurement by identifying a set of variables or factors influencing risk. This work was done very much in the context of the theory of portfolio selection, and as such attempts to model aspects of the economic environment in which share returns are generated. It is assumed that a decision maker can identify the specific factors that are relevant. Relationships between share returns and the factors are then represented by response coefficients which become surrogate risk measures, so that the unidimensional variance measure is replaced with a truly multidimensional reflection of risk. The question is whether the conceptual multidimensional risk framework can be generalised in the context of MCDM, and whether it is practical and tractable to do so. It is important to note that within the financial environment, risk is still focused on only one criterion, the variability of share returns. In the more general problem of MCDM under uncertainty, there are several criteria that must be evaluated by treating uncertain outcomes. In this context a multiple factor model is needed to characterise the risk underlying each criterion, which may be overly time consuming if possible at all. However, in mitigation of these setbacks, if the aim of the decision process is taken to be a better understanding of the problem and potential solutions, then a full exploration into the components of uncertainty may well illuminate some issues for the decision maker.

From the perspective of general MCDM, two issues are of special importance. Firstly, the research into risk-value models has been focused on the case of a single attribute,

usually share returns. Although the theoretical modelling of risk has progressed to a great degree since the developments of Markowitz [59], the challenge to MCDM is to incorporate the tradeoff aspects of the multiple criteria model with the single criterion risk-value model. In particular, it seems necessary to provide a framework to explicitly trade off risk in one attribute with risk in another. The basic structure of such a framework might follow closely the conventional structure of value tradeoffs, although it would seem almost certain that compensations would be far more difficult to interpret in a risk setting, perhaps necessitating some modifications. Second, the more sophisticated measures of risk within the framework of the risk-value models are dependent on the specific form of the utility function, so that at least some modelling of the utility function is necessary before the risk measure is available. Bearing in mind that the reason that explicit risk measures were considered was to avoid the specification of the utility function, and the ensuing complication in the multiple criteria aspects of the problem, their use appears for the time being restricted towards a less facilitated 'backroom' analysis.

3.5 Summary of Research Questions

As a final point to conclude this chapter and bridge into the next, we attempt to summarise and categorise the main research questions which are raised by the discussions in this chapter. Although the categories are rough and not mutually exclusive, three investigative regions can be identified. Questions of behaviour refer predominantly to the abilities of DM's to operate in the complex and taxing decision environment. Generally, they address concerns that too many methodologies do not give enough standing to softer procedural issues, so that models appear technically sound but difficult to implement. Questions of methodological development consider steps in the construction of a decision model that have either been passed over by the literature or not located. Then, it has been a common concern that different MCDM methodologies may provide strongly different results to the same decision problem. A common and intuitively attractive justification is that the different methodologies emphasise different aspects of the problem, so that the same results cannot be expected. However, where different versions of the same methodology exist, as in the outranking and GP methods, we may expect a greater degree of consistency. Finally, questions of pragmatic simplifications may be asked in an attempt to find a balance between realism, ease of use and accuracy of results. We may search for the aspects of a decision problem which materially affect results – and in doing so find a method that excludes unnecessary detail without neglecting that which is important.

Questions of Behaviour

1. When a value-focused approach to scenario development is used, do DM's become aware of changing preference structures over the different scenarios?
2. Are DM's able to think and isolate their preferences conditional on specific

scenarios? That is, is it feasible to elicit preference information in different scenarios? To what extent does anchoring to the current scenario occur?

3. Are DM's able to understand and use the stochastic outranking methods? Is the level of sophistication too high? How comfortable are DM's with specifying thresholds in terms of probabilities?
4. To what extent do heuristics and biases known to occur during the assessment and interpretation of probabilities affect different MCDM methodologies?
5. Are DM's able to specify aspirations in terms of probabilities? Is the revision of aspirations more difficult in a probabilistic environment?
6. Are DM's more comfortable specifying weights on criteria and scenarios simultaneously or separately?

Questions of Methodological Development

1. How should the meta-aggregation of results in each of the p scenarios proceed? How should the notion of robustness be operationalised? How should the notion of approximate robustness be operationalised?
2. How similar are the rank orders produced by the different stochastic outranking methods? Do they provide different degrees of decision aid in terms of opportunities to learn, ease of use and interpretability?
3. How should probabilities be incorporated into the discordance measures of the scenario-based outranking methods?
4. How similar are the rank orders produced by the different stochastic GP methods? Do they provide different degrees of decision aid in terms of opportunities to learn, ease of use and interpretability?
5. What kind of metrics are compatible in the criterion-wise and scenario-wise aggregation stages of the scenario-based metric model? Should different metrics be allowed at all?

Questions of Pragmatic Simplifications

1. To what extent does the use of MAUT affect results when preferences change over scenarios?
2. To what extent do the results obtained by MAUT and non-EU models differ in a practical environment where the axioms of EUT are violated?
3. To what extent do stochastic attribute values affect the results of a decision model? Is the 'ignoring' of uncertainty by estimating expected or likely attribute values a viable option?

4. Do non-idealities such as elicitation errors and independence violations, known to occur in practice, significantly affect the results of an MCDM method under conditions of risk?
5. What scenario selection policies are appropriate? To what extent do non-idealities impact on the selection? How many scenarios should be chosen? Should probabilities be assigned to the scenarios?
6. Are the considered methodologies affected differently by the condition of risk? That is, are the results of certain methodologies more or less sensitive to randomness in the attribute values? Do non-idealities differentially affect the methodologies? Are scenario selection policies consistent over the methodologies?

University of Cape Town

Chapter 4

Simulation Studies in MCDM

4.1 Research Experiments in MCDM

Experiments designed to respond to research questions in MCDM have generally taken the form of observations of real-world MCDM applications, observations of student groups in a controlled environment, or simulation studies of conjectured problems. Within MCDM in particular, but applicable to most modelling disciplines, real-world experiments are severely restricted by the small number of cases that may be considered. Each decision process requires considerable resources, time and effort, so that decision makers are unlikely to be prepared to participate in academic comparisons of multiple decision models. Furthermore, it is far more difficult to perceive the dimensions on which each real-world decision problem is different, so that it is difficult or impossible to isolate the marginal effects of the aspects under study. This, coupled with the small number of considered cases, makes general inferences more difficult.

However, where the research questions relate to the behavioural questions identified previously, a real-world experiment is in most circumstances the only approach capable of offering an answer. Within the context of this study, a simulation approach was chosen on the basis of the time and cost restrictions outlined above. This necessarily excludes from the focus of the experiments those aspects that are behavioural in nature. The remaining questions are all interesting and potentially important; some narrowing of focus is required, however. For the purposes of the simulations which form the remainder of this thesis, we select the questions which appear to probe the groundwork of risk treatment in MCDM. We select the following three questions for closer consideration:

1. Can the 'ignoring' of uncertainty by simplification strategies be justified?
2. Do non-idealities known to occur in practice significantly affect the results under conditions of risk?
3. In applying a scenario-based MCDM approach, to what extent do results dif-

fer depending on how many scenarios are chosen and what scenario selection policies are employed?

These three questions are obviously interrelated: whether certain simplification or selection strategies are appropriate may conceivably depend on whether certain non-idealities are present. The questions cannot therefore be answered in a strict linear fashion, but as a whole they form a strong set of foundations for any future work. Although the questions are all relevant for any of the MCDM methodologies presented previously, aiming for succinctness we attempt to find answers only for the value function methods. The simulations described for the GP and outranking methods are limited to proposals for future research.

4.2 The General Simulation Structure

All simulation studies in MCDM are faced with the difficult question of which aspects of the decision process to include as parameters of the model. Some of the detail of the real-world decision process should be excluded without oversimplifying the model to the point where the results are not clearly interpretable in the real world. In short we want to abstract the model to a point where the results obtained are easily interpretable yet capable of making meaningful recommendations for real-world decision processes.

The starting point of the simulation study is the assumption that the DM possesses an idealised underlying preference structure, which exists as a goal which is aimed towards even if its actual form is not consciously known. This idealised preference structure is here assumed to satisfy the properties of completeness, transitivity, consistency, continuity, and mutual utility independence i.e. preferences may be represented by multiplicative utility functions. This conjecture should not be associated with the (behavioural) idea that underlying every DM is a multiplicative utility function. Rather it is an explication of the guiding nature of MCDM methods in the construction of a preference structure which accurately represents the DM, and may be comfortably defended. Most importantly, the idealised preference structure results in a 'true' rank ordering, which allows for the comparison of the rank ordering produced by this so-called true preference structure with rank orderings produced by applying various MCDM approaches.

The simulation can be thought of as being composed of five stages:

1. The generation of a problem context external to the preferences of the decision maker.
2. The generation of an idealised preference structure and a consequent true rank ordering.

3. The modelling of certain non-idealities relating to errors introduced into the decision process.
4. The modelling of potential treatments of the uncertain outcomes characterising the decision problem.
5. The comparison of the rank orderings produced by the true and non-idealised simulations.

We consider each of these aspects in more detail in the following sections.

4.3 Implementation of a Simulated Value Function Method

4.3.1 The Problem Context

The problem context is represented by a set of n alternatives evaluated over m attributes for each of p scenarios, and a set of probabilities $\Pr(k)$ on the set of scenarios $k = \{1, 2, \dots, p\}$. The assumption is thus that the true underlying distribution of outcomes for each alternative is discrete. The assumption is made for convenience, but is not particularly restrictive in that for sufficiently large p , continuous distributions can be approximated to any desired degree of accuracy.

The Attribute Evaluations

The evaluations z_{ijk} , representing the performance of alternative i on criterion j when scenario k occurs, will in practice occur almost exclusively in a highly complex manner. For the purposes of a simulation experiment it is necessary to limit the complexity of the governing process to the point where a meaningful interpretation of the simplified structure is possible, without compromising the non-trivial aspects of the process. As a result many such simplified structures can be envisaged depending on the salient aspects that are being investigated via the simulation experiment. At no stage is the simulated structure intended as an even tentative description of reality.

Here, the evaluations are generated according to the process

$$z_{ijk} = A_{ij} \cdot C_{jk} + B_{ik} \quad (4.1)$$

The A_{ij} are intended to model the conjecture that each alternative will tend to identify with certain criteria, in that they perform strongly on some of the criteria in most or all scenarios, and weakly in some criteria. They are therefore interpreted as baseline performance evaluations for alternative i on criterion j when *all* scenarios are considered. Of course some alternatives may perform moderately on all criteria, but this remains a case of an identification with a general (moderate) level of performance

on each criterion. The A_{ij} are generated over the sets of alternatives and criteria, independent of scenarios, according to the following process:

1. Generate each A_{ij} as the square of a normally distributed random variable with a mean of 0 and a standard deviation of 0.1.
2. For each alternative i , standardise the A_{ij} to sum to unity i.e. $A_{i1} + A_{i2} + \dots + A_{im} = 1, \forall i$. We discuss later how this construction ensures that the resulting alternatives are non-dominated in at least one scenario.

The form of the random variable as well as the values for the mean and standard deviation were selected after a number of trial-and-error experiments. Several other forms were considered, such as generating the A_{ij} from a uniform distribution and standardising them to lie on the unit hypersphere, as done in [85]. The form finally selected seemed to provide a balanced but reasonably diverse set of alternatives. For the problem context consisting of 7 criteria, most of the A_{ij} lie between 0 and 0.4.

Naturally there must be some variation of performance over scenarios for the uncertain case to be of any interest. A scenario-criterion interaction C_{jk} may be envisaged whereby certain scenarios have a multiplier effect on the baseline evaluations of criterion j . The motivation for the use of a multiplicative modification is that some scenarios may scale up the magnitude of the evaluations in an environment where a consequential scaling of variation is implied. For example, we may consider the case of projected future costs and earnings dependent on various inflation scenarios. Various magnitudes of scenario-dependent deviations can be simulated by adjusting the ranges of the possible values for the C_{jk} . To be consistent with the interpretation of the baseline evaluations A_{ij} , the C_{jk} must have a geometric mean of one for each criterion. This ensures that the ordering implied by the baseline evaluations is preserved in an average sense. It is worth mentioning that since the C_{jk} are independent of the set of alternatives, they do not directly impact any relative rank orderings. Their inclusion is largely due to the intuitive attractiveness of the resultant model and their indirect influence on the implied relative scaling of the additive modification B_{ik} .

The B_{ik} are deviations from the baseline evaluations of alternative i resulting from different levels of performance dependent on scenario k . Thus various magnitudes of scenario-dependent deviations can be simulated using different ranges for the B_{ik} . This additive perturbation is motivated by the conjecture that some alternatives identify to a greater or lesser extent with certain scenarios, which serves to increase or decrease the baseline performance evaluation. This difference is applied equally to all criteria, although the different magnitudes of the C_{jk} ensure that the relative magnitudes of the differences are quite different. Of course the influence of the scenarios is unlikely to be as simple in practice; a specific scenario may increase performance

on some criteria and decrease it on others. The B_{ik} have been generated to have an arithmetic mean of zero, for each alternative, so that the order implied by the evaluations A_{ij} are on average preserved for each criterion. This restriction guarantees that there is no shift in baseline performance due, for example, to positive modifications on all scenarios. Together the B_{ik} and C_{jk} represent the extent to which the problem context i.e. the attribute values, changes over the set of scenarios. The z_{ijk} are then standardised within each criterion to lie in the interval $[0,1]$. Finally, in order for (4.1) to be mathematically sound the three two-dimensional variables are each held constant over their respective third variables in the creation of the three-dimensional z_{ijk} .

There are two points requiring clarification. Firstly, the first scenario has been generated as a scenario where the average rules, and is termed the ‘status quo’ scenario. This implies that $B_{i1} = 0, \forall i$ and $C_{j1} = 1, \forall j$. The assumption is motivated and justified by the reality that scenarios are often constructed so as to place the status quo between some best- and worst-case scenarios. Although in some cases the real status quo scenario may be fairly extreme, it might be expected to be fairly central in the long run. In any case it provides a useful starting point for the comparison and ultimately selection of other scenarios. Secondly, dominated alternatives should be discriminated against by any MCDM methodology, and in a simulation context serve only to inflate the reported capabilities. It is therefore desirable to work with a non-dominated set of alternatives. In the considered context of scenarios, there are several levels of non-dominance. Firstly, there is what could be termed total non-dominance, where each alternative is non-dominated in every scenario i.e. for any two alternatives a and $b, \forall k : z_{ajk} \geq z_{bjk}, \forall j$, with at least one strict inequality. For the purposes of this simulation, such a condition is considered too harsh and restrictive. Instead, we require only that alternatives are non-dominated on at least one scenario. This was achieved by restricting the A_{ij} to sum to one so that, within the context of the status quo scenario, all alternatives are non-dominated. It will usually be true that non-dominance holds on a far larger set of scenarios than just the status quo.

The Scenario Probabilities

The relative scenario probabilities were generated uniformly between 0.2 and 1 and standardised to sum to unity. For $p = 50$ scenarios, this meant that the probabilities generally lay between 0.007 and 0.035. Within a single replication, the difference in relative likelihood between the most- and least-likely scenarios is limited from above by a factor of 5, but is less than 3 in only 5% of all replications.

4.3.2 The Idealised Preference Structure

The idealised preference structure consists of the marginal utility functions used in the construction of the global preference function, and the criterion weights.

The Utility Functions

The utility functions used in the construction of the idealised preference structure are based upon the characteristics of diminishing sensitivity i.e. risk proneness for losses and risk aversion for gains relative to a reference level, and loss aversion [95]. This implies a utility function which is convex below a reference level and concave above it, and which is steeper below the reference point. Each marginal utility function is fully described by four parameters, following the ideas of Stewart [86]: the reference level, τ_j , the value of the utility function at the reference level, λ_j , the curvature of the utility function below the reference level, α_j , and the curvature of the utility function above the reference level, β_j , and is of the standardised exponential form

$$u_j(x) = \begin{cases} \lambda_j(e^{\alpha_j x} - 1)/(e^{\alpha_j \tau_j} - 1) & \text{for } 0 \leq x \leq \tau_j \\ \lambda_j + [(1 - \lambda_j)(1 - e^{-\beta_j(x - \tau_j)})]/(1 - e^{-\beta_j(1 - \tau_j)}) & \text{for } \tau_j \leq x \leq 1 \end{cases} \quad (4.2)$$

Quite a diverse set of preference types may be simulated by adjusting values for τ_j and λ_j . The parameter λ_j is an indication of the strength of preference for avoiding performances below the reference level τ_j for criterion j , so that the severity of the preference threshold separating losses and gains increases in λ_j . Further details are outlined in section 4.4. The use of a single utility function for each criterion together with the [0,1] standardisation of the z_{ijk} within each criterion implies the MAUT feature that although some scenarios may occupy different segments of the marginal utility function, preferences are constant over scenarios (see section 3.1.2, page 37 for a discussion).

The Criterion Weights

The criterion weights are defined in an average sense, generated uniformly on the interval [0.1, 1] and standardised to sum to unity. For $j = 7$ criteria, this ensures that the magnitude by which one criteria weight may exceed another is limited to a factor of 10, and will in most circumstances be considerably lower (a maximum factor of less than 6 is reported in almost 70% of the simulations). The weights are held constant over scenarios.

Construction of the True Rank Ordering

The global utilities are determined in a multiplicative model

$$1 + \kappa U(a) = \prod_{j=1}^m [1 + \kappa w_j u_j(a)] \quad (4.3)$$

where κ is the multivariate risk proneness coefficient. Values of κ greater than zero correspond to multivariate risk proneness, while $\kappa \in (-1, 0)$ indicates multivariate risk aversion. The final calculation of the utilities is performed by taking expectations over the p scenarios, which produces a complete order, termed the true rank order.

4.3.3 The Modelled Non-Idealities

Additive Representation of Utilities

Previous studies by Stewart [85] discussed in subsection 3.1 showed that results were relatively insensitive to approximation of the multiplicative global utility function by an additive form. In light of these findings, and noting the considerable resulting conceptual simplifications, additive utility functions are used throughout the simulations of the non-idealities. The results have been segmented so as to provide evidence of the impact of the additive approximation relative to the other simulated non-idealities.

Omission of Criteria

Although the omission of criteria is a structural issue that may therefore befall any MCDM methodology, the sensitivity of particular methods to the omission of criteria is investigated as part of the robustness study. The number of omitted criteria, o , is a parameter of the simulation. We conjecture again that it is likely that the most unimportant criteria are omitted from the analysis, but that there is some scope for error in this regard. The $m - o$ criteria are therefore selected with probabilities proportional to the ranks of their importance weights, and weights on the selected criteria are standardised once more to sum to one.

Violations of Preferential Independence

The idealised preference structure constructed in section 4.3.2 represents a utopian decision making context, with one consequence being that the criteria satisfy at least the condition of mutual utility independence required for the existence of a multiplicative global utility function. However, the real-world decision process may result in the elicitation of criteria that violate such independence assumptions by being utility dependent. A simple but effective way used by Stewart in [86] and [90] to model these independence violations is to consider mixtures of the idealised criteria, which are by definition independent. For each pair of idealised, utility independent criteria (z_j and z_{j+1} , for $j = \{1, 3, 5, \dots\}$), we can produce two criteria violating utility independence (\tilde{z}_j and \tilde{z}_{j+1} , for $j = \{1, 3, 5, \dots\}$) using

$$\tilde{z}_j = (1 - \gamma)z_j + \gamma z_{j+1} \quad (4.4)$$

$$\tilde{z}_{j+1} = \gamma z_j + (1 - \gamma)z_{j+1} \quad (4.5)$$

where γ is the mixing parameter governing the degree to which utility independence is violated. It may take on values in the interval $(0, 1)$, with maximum violations of utility independence occurring when $\gamma = 0.5$.

Shifts in Reference Levels

The degree of approximation of the elicited reference level to the idealised reference level is also dependent on the experience and conceptual abilities of the DM, and

may therefore be subject to errors or shifts either side of the ideal. The shift itself is viewed as a definitive misinterpretation on the part of the decision maker, and is consequently modelled using a fixed value. The size of the shift is a parameter of the simulation, e_r , which may take on positive or negative values. Its impact is in a sense to blur the target represented by the idealised preference structure by changing the shape of the marginal utility function before it is approximated piecewise linearly in the elicitation process.

Estimation of utility functions by v piecewise linear segments

In practice the idealised marginal utility functions are elicited by a series of questions resulting in a piecewise linear utility function approximating the idealised utility function. The number of segments in the piecewise linear function, v , reflects the time and effort involved in the elicitation, and is another parameter of the simulation. For each modelled criterion \tilde{z}_j defined previously, $v + 1$ equally spaced points are chosen along the attribute axis from the minimum to the maximum values. The value of the marginal utility function at each point is defined and linearly interpolated between. In cases where some violations of independence occur i.e. $\gamma \neq 0$, the attribute values will not in general cover the full interval $[0, 1]$, so that the resulting piecewise linear utility function must be standardised to the interval $[0, 1]$. In such cases the weights must also be modified to reflect the differences in the relative attribute ranges over which the idealised swing weights were originally defined. This is done using proportions between the minima and maxima of the non-idealised utilities on each criterion. Once all restandardisation is complete, the new piecewise utility functions are used to read off the appropriate non-idealised marginal utilities.

Errors in Weight Assessment

The elicitation of criterion weights from the DM is a well-known area for the potential introduction of errors into the decision process, particularly where the DM is unfamiliar with the elicitation process or has poorly formed preferences. Either an additive or multiplicative form for the errors can be envisaged; it should not be materially important which is used. While acknowledging that a multiplicative form may be justified in cases where the criterion weight is so close to zero that additive errors may result in non-negative weights, we make the conjecture that decision makers view criterion weights as fuzzy but falling somewhere within intervals of reasonably consistent width, where that width is determined by the confidence and knowledge of the DM. The additive assessment errors are therefore generated uniformly on the interval $[-e_w, e_w]$, where e_w is a parameter of the simulation controlling the size of the errors. In order to retain the property that the weights sum to one, the average component is also constrained to sum to zero, and restrictions are applied so that criterion weights may not be negative as a result of excessive errors.

Errors in Probability Assessment

The errors in the assessment of scenario probabilities is motivated by much the same reasons as the errors in weight assessments, specifically that uncertainty and unfamiliarity in the decision process may induce errors. The same question of additive or multiplicative representations follows. However, in the case of probability assessment errors, the small magnitude of the probabilities made assigning additive errors more difficult. Multiplicative errors are therefore used, generated uniformly on the interval $[1 - e_{Pr}, 1 + e_{Pr}]$, where e_{Pr} is a parameter of the simulation controlling the size of the errors. The resulting non-idealised probabilities are standardised once more to sum to one.

4.3.4 Treatment of Uncertain Outcomes

The non-idealities raised in the previous subsection exist independently of the uncertainty aspect of the problem – they are equally as valid when outcomes are deterministic and certain. When outcomes are uncertain, it remains a fundamental choice of the decision analysis how to pursue this uncertainty. We may attempt to model it in its entirety, break it into ‘manageable pieces’, or ignore it entirely. This subsection attempts to incorporate different elements of these choices into the simulation environment constructed thus far.

Ignoring the Risk

However unsophisticated ignoring the uncertain aspects of a decision problem may sound, there may be adequate justification for it if time and resources are limited, or if a low robustness is permissible. Further, the impact of ignoring at least some uncertainty is an important question in its own right. It has already been argued that it is not so much whether uncertainty is to be ignored, but to what extent. In this subsection we consider two ways to model a plausible heuristic that transfers the decision problem from a stochastic to a deterministic framework. For simplicity in both cases the non-idealised value functions take the same form as the non-idealised utility functions, although in practical cases the different elicitation techniques are likely to cause differences in the resulting functional forms.

A simple way in which to model a simplification strategy is based on the use of expected attribute values. In circumstances where the DM wishes to avoid the complexity of a stochastic MCDM method, an attempt may be made to compute the expected values of each z_{ij} and use those values as inputs to a deterministic model. The two models developed here are based on this view. Naturally the success of this simplifying strategy will depend on the care taken in the construction of the expected attribute values. At the most conscientious level, the DM might make explicit the attribute values z_{ijk} in each of the p scenarios as well as the scenario probabilities $\Pr(k)$. Although this would seemingly negate most of the time- and effort-saving qualities

of the strategy, such an approach might be justified based on the easier elicitation of marginal value functions. In any case, this represents the best that the DM might do given the non-idealised information. On the other hand, the DM might simply perform the analysis after assuming that one of the scenarios is the deterministic future, discarding all other potential scenarios from the analysis.

We may therefore use the approximation to $E[z_{ij}]$ as a model of the conscientiousness of the DM, using convex combinations of $E[z_{ij}]$ and z_{ijS^*} , where S^* is the most likely scenario. Formally,

$$z_{ij}^{EZ} = \Lambda E[z_{ij}] + (1 - \Lambda)z_{ijS^*} \quad (4.6)$$

where Λ is a parameter increasing from 0 to 1 with the conscientiousness of the DM in obtaining the expected attribute values. The parameter $1 - \Lambda$ can also be considered as the degree of anchoring to the most likely scenario in the estimation of expected attribute values. For the purposes of these simulations we consider the cases of $\Lambda = 0.2$ and $\Lambda = 0.8$. We refer to the first model as ‘ignoring risk’ and the second as the ‘expected Z’ model. Finally, it should be emphasised that no claim is made that DM’s think strictly in terms of convex combinations or even that they consider only the two quantities outlined above; the models above merely provide a mechanism for the examination of what is in all likelihood a highly complex process.

Selection of $s < p$ scenarios for analysis

If p is even relatively large, the evaluation of attribute values and DM preferences can quickly become unwieldy, and a natural question is to what extent the selection of $s < p$ ‘representative’ scenarios affects the rank ordering produced by the analysis. Two fundamental questions of this thesis relate to how best to select ‘representative’ scenarios and what number of scenarios should be selected to ensure a good approximation. In this regard, the number of selected scenarios s is a parameter of the simulation, and four selection policies are evaluated:

1. Scenarios are selected randomly.
2. Scenarios are selected with probabilities proportional to their respective probabilities of occurrence.
3. Selection of the s most likely scenarios.
4. Selection of the status quo scenario and $s - 1$ maximally diverse scenarios as determined by a forward filtering algorithm.

Selection policy 1 (random selection) represents a worst-case scenario in which the best that can be achieved is a random choice of s scenarios. The DM may be unable to provide any information as to the relative likelihoods of the

scenarios, or may be unwilling to use them as a basis for selection. This policy may be appropriate when extremely little qualitative information is available on the nature of the scenarios i.e. the scenarios are considered to be nothing more than realisations of an unknown multivariate random variable.

Selection policy 2 (proportional selection) characterises a policy that acknowledges the importance of relative likelihoods in the selection process, but is subject to fairly large errors in this selection process. The corrupted scenario probabilities are used in the selection, and although these probabilities are assumed known to the DM and are used in the construction of the final scores, there remains some scope for other policies and other errors. The actual nature of these problems is not overly important, but may include a reluctance to rely only on probabilities, a piece of qualitative information, or a subjective ‘hunch’. Therefore although the most likely scenario is the most probable initial selection, the actual selection remains random. Given that the probabilities of occurrence are reasonably close together, the scope for error in the attempted probabilistic selection is substantial, and the process should probably be considered closer to purely random than selection of the s most likely scenarios.

Selection policy 3 (most likely scenario selection) represents the logical end-point of selection policy 2 – the selection of the s most likely scenarios. This is a deterministic selection policy that hypothesises that the DM has available the corrupted scenario probabilities and wishes to use them exclusively. The question of interest is to what extent the tactic of purely probabilistic selection is a valid and wise choice. Selection policies 1-3 provide a graded set of results intended to convey the importance of selecting scenarios on the basis of relative likelihood (which is not advocated by scenario planning experts [96]).

Selection policy 4 (filtered selection) better conveys both the advice of scenario planners and the heuristic which we conjecture is often used by decision makers in practice. That is, starting from a status quo position, which we have defined as the first scenario i.e. $k = 1$, the DM selects $s - 1$ maximally different scenarios. This is most usually observed as a ‘bad-average-good’ design when $s = 3$, but is easily extended to incorporate more scenarios. Section 3.1.2 discusses selection policies in more detail. The selection of divergent scenarios is likely to be a complex process with considerable scope for serious error; for the purposes of this simulation, the selection is modelled using a forward filtering algorithm to select scenarios which are in a sense most different to the status quo.

The Forward Filtering Algorithm The forward filtering algorithm employed here uses the method of furthest point outside the neighbourhood [84]. For the purposes of the algorithm each scenario is considered to be represented by a column vector in the \mathbf{B} matrix i.e. by a point in p -dimensional space. Taking as the starting point the status quo scenario i.e. $B_{i1} = 0, \forall i$, a hypersphere of radius r (a parameter of the algorithm, see [84] for more details) is constructed

around the point representing performance on the status quo scenario i.e. the origin. The scenario possessing the greatest Euclidean distance between itself and the status quo scenario is selected as the second scenario, provided that it does not lie within the hypersphere. A similar hypersphere is constructed around the second point, and the scenario possessing the greatest Euclidean distance between itself and the second chosen scenario is selected as the third scenario, provided that it does not lie within either the original or the second hypersphere. The procedure continues until no more scenarios can be selected i.e. all unselected scenarios lie within the radius of one of the hyperspheres surrounding the selected scenarios. If the number of selected scenarios is equal to s , the algorithm stops, otherwise the original radius r is either increased (if the number of selected scenarios exceed s) or reduced (if the number of selected scenarios is less than s). This process continues iteratively until the number of scenarios selected by the forward filtering algorithm equals precisely s .

4.3.5 Comparisons of Rank Orders

Construction of Model Rank Orders

For each set of non-idealities considered, eight rank orders are produced corresponding to

- The additive model with no reduction of uncertainty i.e. all scenarios considered.
- The model carelessly considering expected attribute values i.e. ignoring risk.
- The model diligently considering expected attribute values.
- The random scenario selection policy.
- The proportional scenario selection policy.
- The most likely scenarios selection policy.
- The filtered selection policy.

Measures of Comparison

The results of the simulations are presented in the form of the following four measures.

1. POSN. The average position of the true best alternative (according to the idealised rank ordering) in the model rank order.
2. RANK. The average rank of the best alternative selected by the model in the true rank order.
3. Spearman's Rank Correlation coefficient calculated between the model rank order and the true rank order.

4. A rank probability matrix \mathbf{P} with element p_{qr} denoting the probability that the alternative with rank q in the true rank order has rank r in the model rank order. The rank probability matrix varies between the identity matrix in cases of complete agreement and a matrix having all elements equal to $1/n$ in cases of a complete lack of agreement.

4.4 Parameters of the Simulation Experiment

All value function simulations were based on a problem containing $n = 15$ alternatives, $m = 7$ criteria and $p = 50$ scenarios, representing as best as can be expected a ‘typical’ MCDM problem. In the long run we could therefore expect a worst-case model randomly selecting a ‘best’ alternative to obtain POSN and RANK scores of 8.

Changing attribute values over scenarios

Changes to the attribute values over scenarios were simulated using the additive and multiplicative scenario-based modifications, B_{ik} and C_{jk} respectively. Larger values for B_{ik} relative to C_{jk} imply greater change in attribute values over the range of scenarios and a more dominant scenario effect. The values of B_{ik} and C_{jk} were adjusted with respect to the baseline score A_{ij} , where the A_{ij} were generated so that approximately 90% were less than 0.4. Since it is only the relative scaling of the B_{ik} to the C_{jk} that materially influences the attribute values, only one magnitude of multiplicative changes, C_{jk} , was used, effectively lying between 80% and 120%.

Parameter	low	high
B	U(-0.2, 0.2)	U(-0.3, 0.3)

Table 4.1: Simulated levels of attribute variability

Basic preference structure

Four different basic preference structures were simulated in the manner of Stewart [86], based on the following modifications to the parameters, used in the definition of the utility functions, governing the reference level, τ_j , and the value of the utility function at the reference level, λ_j .

Case	τ_j	λ_j	Description
1	U(0.15, 0.4)	U(0.15, 0.4)	Mainly compensatory preferences
2	U(0.15, 0.4)	U(0.6, 0.85)	Mainly compensatory preferences, with a sharp preference threshold
3	U(0.6, 0.85)	U(0.15, 0.4)	Limited compensation
4	U(0.6, 0.85)	U(0.6, 0.85)	Limited compensation, with a sharp preference threshold

Table 4.2: Parameters of the four simulated preference structures

In each case, the curvature of the utility function above the reference level, β_j , was generated uniformly on the interval $[1, 8]$, and the curvature below the reference level, α_j , was generated uniformly on the interval $\beta_j + [2, 8]$. The effect of multivariate risk proneness is investigated in the first experiment by changing the value of the parameter $\kappa = \{1, 3, 8\}$. In the subsequent experiments $\kappa = 3$. The two degrees of change to the attribute values together with the four basic preference structures define the 8 basic problem structures for the simulated implementation of a scenario-based MCDM approach.

Simulation of Non-idealities

The simulation is defined by the following model parameters representing the non-idealities of the decision process. The reference levels of each parameters have been highlighted in bold face.

1. Number of scenarios selected for analysis. $s = \{3, \mathbf{5}, 10\}$
2. Errors in weight assessment. $e_w = \{\mathbf{0}, U(\pm 0.05), U(\pm 0.10)\}$
3. Errors in probability assessment. $e_{Pr} = \{\mathbf{0}, U(\pm 0.3), U(\pm 0.6)\}$
4. Estimation of utility functions with piecewise linear segments. $v = \{1, 2, 4\}$
5. Omission of criteria. $o = \{\mathbf{0}, 1, 2\}$
6. Violations of preferential independence. $\gamma = \{\mathbf{0}, 0.1, 0.3\}$
7. Shifts in reference levels. $e_r = \{\mathbf{0}, 0.1\}$

Investigation of a scenario-based value function method

Three value function experiments have been performed in order to keep computational demands manageable. The first experiment isolates for closer investigation the scenario issues constituting our core concern. Then, the second and third experiments investigate the effects of certain non-idealities known to occur in practice, with a specific view to investigating further any interactions between the non-idealities and the use of a scenario-based approach.

Experiment 1 Simulation of non-ideality 1 and the multivariate risk proneness parameter κ . All other non-idealities are set at their reference levels, giving a total of 72 different cases.

Experiment 2 Simulation of non-idealities 1, 2 and 3, with all other non-idealities set to their reference levels. For the purposes of this experiment, the set of basic structural parameters governing the shape of the utility function is fixed at case 1, so that there are only 2 basic problem structures. All combinations of non-idealities 1,2 and 3 are simulated, giving a total of 54 different cases.

Experiment 3 Simulation of the non-idealities 1, 4, 5, 6 and 7 previously investigated by Stewart [86]. All other non-idealities are set to their reference levels, the shape of the utility function is fixed at case 1 and attribute variability is high. All combinations were used to give a total of 162 different cases.

The results of these three experiments are presented and discussed in chapter 5. In the remainder of the current chapter we propose hypothetical simulation structures for similar investigations in the context of each of the outranking and metric methodologies.

4.5 Proposed Implementation of a Simulated Metric Method

Sections 4.3 and 4.4 gave a complete and detailed account of a value function simulation approach, as it was implemented for the purposes of this thesis and for which results appear in chapter 5. On the other hand, we did not implement the following proposed simulation of a metric methodology. The discussion that follows is therefore intended purely as an introductory and incomplete outline of the form that such a simulation approach might take, with the specific aim of highlighting any potentially troublesome modelling issues.

In simulating a metric method the basic five-stage simulation structure would remain the same, and it is only the modelling of the MCDM treatment of the problem (stage 4) that would need to change. The problem context, idealised and non-idealised preference structures, and comparison of true and modelled rank orders would all be generated as before. However, whereas in the previous simulation we needed only to consider the single value theoretic model, there is no such consensus of opinion in the metric methods. We have identified three distinct metric methodologies in section 3.3, which we refer to as chance-constrained GP, mean-variance GP and scenario-based GP. The simulation framework therefore would need to consider an additional research question

- How similar are the rank orders produced by the different stochastic GP methods?

In retaining the general simulation structure, the simulated DM is again assumed to possess a specific preference structure that is consistent with the ‘rationality’ properties of expected utility theory. This idealised preference structure is not directly available to the DM, but is in some sense strived towards through the haze of simulated non-idealities. This is particularly important in the context of metric methods, where an unwillingness to confront the difficulty of eliciting utilities may often motivate their use. This formulation in no way implies that metric methods are an inferior approximation to results obtained using expected utility theory, or that beneath a metric approach lies a DM thinking in terms of utilities and associated concepts. Rather the axioms of expected utility theory are used in the absence of any other normatively appealing axiomatic basis for the comparison of MCDM methods. In fact these axioms play a limited role in the proposed simulation of metric methods, being used primarily as a basis for an optional and unsophisticated revision of aspirations.

In the remainder of the section we propose a revised fourth stage of the simulation process appropriate for a metric approach. The basic algorithm is as follows:

1. Elicit an initial set of aspirations.
2. Generate the set of deviations.
3. Apply a metric method of aggregating the deviations into a score.
4. Rank order the alternatives and present the solution to the DM.
5. (Optional) If the best alternative has been seen before, then stop. Otherwise, revise the aspirations and return to step 2.

4.5.1 Initial Aspirations

The main focus of the proposed simulation study is not on the revision of aspirations, but rather on how various metric methods accommodate uncertain outcomes in the presence of certain non-idealities known to occur in practice. This focuses attention on the quality of the initial selection of aspirations, since these aspirations will undergo only limited and unsophisticated revision. Further, it has been suggested that well-selected initial aspirations considerably aid the metric process [7]. All aspirations would be set using the non-idealised preference structure.

Scenario-Based GP

Scenario-based GP approaches require that aspirations are set in each criterion-scenario combination. The proposed modelling of the aspiration setting procedure is motivated by two extremes. In cases where there is very little awareness of the DM’s preference structure, aspirations may be simply set to the best performance

i.e. the ideal, in each criterion-scenario combination. At the other end of the scale, where the DM possesses perfect knowledge of all the inputs and products of the idealised decision process (ignoring the question as to why the analysis is then required), aspirations would be set to the performance of the true best alternative on each criterion-scenario combination, to ensure that the best alternative is in fact chosen. The competence of the DM in selecting initial aspirations could therefore be parsimoniously modelled by considering convex combinations of the aspirations corresponding to the respective selection strategies. Let the highest attribute value under criterion j and scenario k be denoted by z_{jk}^* , and let the performance of the true best alternative under criterion j and scenario k be denoted by z_{jk}^{TB} . Then the aspiration g_{jk} for criterion j and scenario k would be given by

$$g_{jk} = \rho z_{jk}^{TB} + (1 - \rho) z_{jk}^* \quad (4.7)$$

where ρ is a parameter increasing with the competence of the DM in setting initial aspirations ($0 \leq \rho \leq 1$).

Mean-Variance GP

The mean-variance GP approach requires that aspiration levels are specified in terms of the expected utility of criterion j rather than in terms of the attribute values. Aspirations g_j are thus set over all scenarios simultaneously. We might use the goals g_{jk} constructed for the scenario-based approach and the non-idealised marginal utility functions to find the utility of each criterion-scenario goal $u_{jk}(g_{jk})$. The scenario probabilities could then be used to form the expectation

$$g_j = \sum_{k=1}^p \Pr[k] u_{jk}(g_{jk}) \quad (4.8)$$

Chance-Constrained GP

As we have mentioned before, CCGP approaches are not well-suited to discrete choice MCDM problems due to the requirement that aspirations be set both on attribute values and the probabilities of achieving these attribute aspirations. The relationship between the two is unclear, although this might conceivably form an additional aspect of the simulation. The attribute aspirations might be specified in each criterion using expectations over scenarios $b_j = \sum_{k=1}^p \Pr[k] g_{jk}$. Alternatively, we may apply the same technique as for the scenario-based approach except that we identify the ideal performance and the performance of the true best alternative in each criterion regardless of scenario. It is difficult to see any specification of the aspirational probabilities g_j apart from an *ad hoc* selection.

4.5.2 Computation of Deviations

In certain generalised goal programming or reference point approaches, the deviations are not constrained to be non-negative, so that surplus performance over and above

the goals is also taken into consideration. For the purposes of this proposed simulation, the more classical notion of non-negative goal programming deviations would be employed.

Scenario-Based GP

The computation of deviations in the scenario-based GP approach would take the form of a trivial step based on the attribute values and the previously generated set of aspirations.

$$\delta_{ijk} = \max[g_{jk} - z_{ijk}, 0] \quad (4.9)$$

Mean-Variance GP

The mean-variance approach does not consider deviations in the usual sense of the metric methods. The condition that $\delta_{ij} = \bar{z}_j - g_j \geq 0$ would create a feasible region and in doing so exclude some alternatives from further consideration. A rank order among the remaining alternatives would then be based on values of the variance-based objective function.

Chance-Constrained GP

To compute the deviational probabilities we would first need to calculate the probability P_{ij} that alternative i exceeds the attribute aspiration b_j for criterion j . In the context of discrete choice, this is easily done by summing together the probabilities of those scenarios for which the condition is true. The deviations from the aspirational probabilities could then be computed as

$$\delta_{ij} = \max[g_j - P_{ij}, 0] \quad (4.10)$$

4.5.3 Aggregation of Deviations

Scenario-Based GP

The scenario-based approach is unique in this context in that it requires two stages of aggregation. We limit consideration to the Archimedean or Tchebycheff aggregations applied at either of the criterion-wise or scenario-wise aggregation stages, giving a total of 4 possible combinations. A detailed motivation for the consideration of these aggregations is discussed in section 3.3.3; here we present only the form of the final score for each alternative i given by the aggregation combination denoted by {criterion-wise aggregation; scenario-wise aggregation}.

{Archimedean; Archimedean}

$$\sum_k \Pr(k) \sum_j w_j \delta_{ijk} \quad (4.11)$$

{Archimedean; Tchebycheff}

$$\max_k \left[\Pr(k) \sum_j w_j \delta_{ijk} \right] \quad (4.12)$$

{Tchebycheff, Archimedean}

$$\sum_k \Pr(k) \left[\max_j [w_j \delta_{ijk}] + \epsilon \sum_j w_j \delta_{ijk} \right] \quad (4.13)$$

{Tchebycheff; Tchebycheff}

$$\max_k \left[\Pr(k) \left[\max_j [w_j \delta_{ijk}] + \epsilon \sum_j w_j \delta_{ijk} \right] \right] \quad (4.14)$$

where ϵ is a small positive constant of order 0.01 used in order to avoid the possibility of non-efficient solutions. Note that the scenario-wise aggregation takes into account the relative likelihoods of each scenario; the advantages of this extension are discussed in section 3.3.3.

Mean-Variance GP

As we have mentioned in the previous section, the mean-variance GP approach does not make use of deviations in the conventional sense; the aggregation stage does not involve deviations at all but rather, for an alternative a , the computation of the variance measure $\sum_{j=1}^m w_j R_j(\bar{z}_j(a)) \sigma^2(z_j(a))$. The value of the local risk aversion at the expected attribute value $\bar{z}_j(a)$ could be calculated from the non-idealised marginal utility functions.

Chance-Constrained GP

The aggregation procedure for the CCGP approach is the most straightforward of the three; because we have a set of m aspirations g_j and a set of m deviations δ_j we may use the conventional GP aggregation making use of a single metric α

$$\left[\sum_{j=1}^m [w_j \delta_{ij}]^\alpha \right]^{\frac{1}{\alpha}} \quad (4.15)$$

4.5.4 Revision of Aspirations

Previous simulations [90] have shown that the revision of aspirations is crucial to the success of a metric MCDM approach. However, we do not pursue in detail the proposed simulation of aspiration revisions for two reasons. Firstly, the model described thus far is quite broad in scope; it would test the effects of GP methodology,

uncertainty, and non-idealities. Furthermore, there may be significant interactions between these effects. The importance of not constructing overly complex or detailed simulation models has already been discussed, and it is with respect to this balance that we choose not to include the considerable further aspect of revising aspirations. Secondly, there has been little axiomatic or empirical work done on how decision makers may adjust their aspirations in light of a particular solution. This complicates the simulation process by necessitating the conjecturing of the process governing aspiration revision. Of course such conjectures, once developed, are useful ways to investigate which aspects of the revision might be more important. These may then form the bases for later empirical studies.

Stewart [90] has proposed a simulated revision procedure in a deterministic GP context. By considering each criterion-scenario combination as a deterministic ‘criterion’, this procedure may be implemented in the scenario-based GP approaches with little modification. The mean-variance approach is more problematic in that revisions would have the effect of reducing or enlarging the set of feasible options; more aggressive revisions might therefore be necessary to induce changes in the alternative selected by the model. However, because the goals of the mean-variance approach are derived directly from those in the scenario-based approaches, it should be possible to again use the procedure proposed by Stewart. The chance-constrained GP approaches present the biggest challenge based on the dual aspirations over attributes and probabilities. Specifically the process by which one might revise the aspirational probabilities is very unclear. One idea is therefore to revise only the attribute aspirations based on the procedure of Stewart, while keeping the aspirational probabilities fixed. Although as a cognitive model this revision is unlikely to be realistic, it would provide a way forward when few appear to exist.

4.6 Proposed Implementation of a Simulated Outranking Method

The simulation of an outranking method is overshadowed by two complications: the lack of an axiomatic basis governing behaviour, and the difficulty in reconciling the only available axiom system i.e. MAUT, with a model designed specifically to avoid complying with those axioms. In particular it is no longer only the fourth stage in the generalised five-stage simulation process that would require modification. The possible incorporation of incomparability implies that the generated model rank order would not necessarily be complete. The comparison of this model rank order to the true rank order might therefore require some modifications. The problem context and non-idealities, however, would be generated as before. The modelling is further complicated by the lack of an accepted method of treating uncertainty. Four broad treatments were identified by the literature survey: direct comparison of probability distributions, construction of a distributive outranking relation, construction of a

stochastic dominance relation, and a scenario-based approach. In this section we again discuss only a *proposed* simulation structure, in order to give a flavour of how an outranking approach might be simulated without going into the amount of detail that would be required for actual implementation. Furthermore we limit our attention to the implementation of a scenario-based simulation approach. In any case, simulations of the other three treatments of risk are likely to be at least similar in nature, differing mainly in the definitions of concordance and discordance. The basic algorithm is as follows:

1. Define the concordance and discordance indices in each criterion-scenario combination.
2. Aggregate the concordance and discordance indices over criterion-scenario combinations to arrive at global concordance and discordance measures.
3. Synthesise the global concordance and discordance measures to construct the outranking relation.
4. Exploit the outranking relation in order to arrive at a (not necessarily complete) rank order.

We discuss first the initial definitions of concordance and discordance in the proposed simulation model, before discussing the final three steps under the heading of aggregation and exploitation.

4.6.1 Concordance and Discordance

The first question that needs to be answered relates to which of the outranking methods to use. The built-in fuzziness of the later ELECTRE and PROMETHEE methods model aspects of imprecision not included in either the value function or proposed metric method simulations. The focus needs to remain on the treatment of risk rather than imprecision. On the other hand it does seem essential not to oversimplify the outranking methods by restricting them to an ELECTRE I or II form. The proposed simulation would make use of the more advanced ELECTRE III methodology, which requires the definition of a set of pseudocriteria using appropriate indifference and preference thresholds. The most important aspect of the proposed simulation is the way that the pseudocriteria are used to form a bridge to the idealised utility functions. Rather than the usual approach that defines the concordance index $c_{jk}(a, b)$ over the domain $z_{jk}(b) - z_{jk}(a)$, we propose to use the domain $u_{jk}(b) - u_{jk}(a)$. This would at least achieve some fidelity to the assumed underlying preference structure.

In order to remain consistent with the other simulations (which did not impose any intransitivity of indifference), it would be necessary to set each indifference threshold t_{jk}^i to zero, although conceivably it too could be left as a model parameter. Thus the concordance index $c_{jk}(a, b)$ would begin to decrease from one as soon as $u_{jk}(b) > u_{jk}(a)$.

The specification of the preference threshold t_{jk}^p is less clear: in the absence of psychological guidelines it seems necessary to experiment with various parameter values in an *ad hoc* fashion. It is important to note that both thresholds must be expressed in terms of utilities rather than attribute values. The proposed concordance index is then given by

$$c_{jk}(a, b) = \begin{cases} 1 & \text{if } u_{jk}(a) \geq u_{jk}(b), \\ \frac{u_{jk}(b) - (u_{jk}(a) + t_{jk}^p)}{-t_{jk}^p} & \text{if } u_{jk}(a) \leq u_{jk}(b) \leq u_{jk}(a) + t_{jk}^p, \\ 0 & \text{if } u_{jk}(a) + t_{jk}^p \leq u_{jk}(b). \end{cases} \quad (4.16)$$

The notion of discordance is more problematic. As it has no analogue in the other MCDM schools it does seem plausible to simply exclude it from consideration, as done in [53]. If it is to be included, due care needs to be taken to ensure that discordances only occur in severe situations. For example, even a difference of $u_{jk}(b) - u_{jk}(a) = 1$ might be considered insufficient to induce discordance unless both the criterion j and scenario k are important. A proposal is then to set the veto threshold t_j^v using a *weighted* difference so that discordance would not occur in less important criterion-scenario combinations. Although this does run contrary to the spirit of discordance, it does not seem unrealistic to expect that similar thinking may occur in practice. A possible discordance index is therefore given by

$$d_{jk}(a, b) = \begin{cases} 0 & \text{if } w_{jk}u_{jk}(b) \leq w_{jk}u_{jk}(a) + t_{jk}^p, \\ \frac{w_{jk}(u_{jk}(b) - u_{jk}(a)) - t_{jk}^p}{t_{jk}^v - t_{jk}^p} & \text{if } w_{jk}u_{jk}(a) + t_{jk}^p \leq w_{jk}u_{jk}(b) \leq w_{jk}u_{jk}(a) + t_{jk}^v, \\ 1 & \text{if } w_{jk}u_{jk}(b) \geq w_{jk}u_{jk}(a) + t_{jk}^v. \end{cases} \quad (4.17)$$

At this point it is probably necessary to warn against the introduction of excessive *ad hoc* simulation elements. The major problem when simulating an outranking approach is the lack of solid physical interpretation for several crucial model aspects, including concordance and discordance. This lack of interpretation makes it considerably more difficult to operationalise the outranking method without giving a very subjective and possibly even arbitrary appearance to the simulation. Some degree of *ad hoc* modelling seems unavoidable, but it is important to limit it as much as possible.

4.6.2 Aggregation and Exploitation

The aggregation of the marginal concordance and discordance indices into global measures to be used as inputs to the so-called credibility index $S(a, b)$ might proceed as outlined in section 3.2.3. The criterion weights present the only difficulty: the different interpretations of the weights in the value function and outranking schools is well known. Previous studies [53] have shown that no clear relationship exists between the criterion weights of the two approaches. However, because the problem context

generates the attribute values so that each criterion is defined over the same unit range, the swing weights of the value function approach have no range component i.e. they reflect purely the esoteric notion of importance. In light of this fact, it seems reasonable to use the same numerical weights in both the idealised preference structure and the outranking model.

Simulating the further aggregation of the ascending and descending distillations resulting from the exploitation of the credibility index is highly problematic, but essential in order to arrive at a final rank order for comparison. The only help in this regard appears to be using a PROMETHEE aggregation procedure, based on the computation of two global scores: a positive outranking flow $Q^+(a)$ indicating the extent to which a outranks all other alternatives, and a negative outranking flow $Q^-(a)$ indicating the extent to which a is outranked by all other alternatives. The flows are defined by

$$Q^+(a) = \sum_{b \neq a} S(a, b) \quad (4.18)$$

$$Q^-(a) = \sum_{b \neq a} S(b, a) \quad (4.19)$$

Each of these flows defines a complete preorder, from which the outranking relation can be obtained in two ways. The PROMETHEE I method uses the intersection of the preorders (see [7]), while the PROMETHEE II method first computes a net outranking flow $Q(a) = Q^+(a) - Q^-(a)$ before concluding that a outranks b if $Q(a) > Q(b)$. The latter aggregation results in a complete preorder, which allows the comparisons of idealised and modelled rank orders to take place unhindered, but is perhaps slightly contrary to the philosophy of outranking as no incomparability is permitted. The former aggregation is more faithful to the outranking philosophy, but would necessitate some difficult modifications to the comparison stage.

Chapter 5

Simulation of Value Function Methods

5.1 Results of Simulation Experiment 1

Simulation experiment 1 is intended to investigate the impact of the more basic aspects of the value function framework, in the absence of other non-idealities. The fundamental features of any value function model are a set of attribute values, a set of marginal utility functions, and an aggregation method governed by a functional form. Accordingly, there are five main effects, two corresponding to the shape of the marginal utility function (τ_j and λ_j), one corresponding to the variability of attribute values over the scenarios (B), one corresponding to the level of multivariate risk proneness (κ) and one corresponding to the number of scenarios chosen (s), taking on the following values to give a total of 72 cases.

Parameter	Values
τ_j	low, high
λ_j	low, high
B	low, high
κ	1, 3, 8
s	3, 5, 10

Table 5.1: Parameters of experiment 1

In this experiment as well as the two experiments that follow it, we begin by briefly discussing an analysis of variance (ANOVA) comprising the main effects and all higher order interactions in the particular experiment. The analysis of variance is used as a quick and concise way to summarise the effects in terms of statistical significance, so as to provide a broad overview of the results obtained. We perform separate ANOVA's for each of the POSN and RANK statistics. In the detailed ANOVA tables that are presented at the end of each of section 5.1, 5.2 and 5.3 it is our convention to identify those effects that are significant at the 5% level by a single asterisk and those effects

that are significant at the 0.5% level by a double asterisk. Following the discussion of the ANOVA results, we consider in more detail the actual nature of each effect and its potential relevance to practical decision making.

5.1.1 ANOVA Results

Detailed ANOVA tables for experiment 1 are presented at the end of section 5.1 in tables 5.12 and 5.13, for the POSN and RANK scores respectively. All main effects are identified as significant at the 0.5% level by the majority of the models. There is a general consensus among the models examined that the reference level exercises the most significant main effect, with F-statistics upwards of 3000. The effect of attribute variability is also identified as highly significant in all cases with the exception of the additive POSN scores. The effect of the number of scenarios chosen will clearly be entirely due to chance in models that operate outside the scenario environment, but within the scenario models it is of a similar magnitude to the effect of attribute variability. The other two main effects, that of the multivariate risk proneness and the utility at the reference level, remain significant at the 0.5% level but are considerably less significant than the other main effects, with F-statistics generally around 100.

There is less consistency between the value function models with regard to the significant interaction effects. There are four second-order interactions that are significant in the majority of the models: between the effect of reference levels and each of the four other effects. Although there are several other interactions identified as statistically significant by some of the models, further investigation into the nature of the interactions revealed little systematic behaviour. These interactions are therefore not of broader practical interest, and are not considered further.

Contrary to the consistency of the main effects and to a lesser degree the second-order effects, the higher-order interactions are usually inconsistent and model-specific where they are significant at all. There is a potentially significant higher order interaction between preferences (jointly represented by τ_j and λ_j) and the effect of attribute variability, which is significant in all models with the exception of the model ignoring risk.

5.1.2 Characterisation of Effects

The general performances of each of the 7 models is reflected in the following table, which shows the grand average POSN and RANK scores as well as the average rank correlations.

Model	POSN	RANK	SRC
Additive	1.40	1.41	0.94
Ignore	5.02	4.33	0.37
Expected Z	2.46	2.21	0.75
Scenario R	4.17	3.96	0.50
Scenario P	4.00	3.81	0.52
Scenario ML	3.46	3.33	0.59
Scenario F	4.36	4.03	0.48

Table 5.2: Grand average performance measures for experiment 1

The additive model, which is the only model that includes a full treatment of the uncertain outcomes, clearly outperforms all the other models, with POSN and RANK scores indicating that the true best alternative is located on average between the first and second model ranks (out of 15 alternatives), and that the alternative identified by the model as best is also between first or second in the true rank order. Although we have not yet incorporated any non-idealities, the results so far mirror those of Stewart [85] to the effect that an additive representation provides an excellent approximation to the true rank order obtained from a multiplicative utility function. The true best alternative is identified as best by the additive model in 75% of the simulations, and among the top two alternatives in 90%. Similar results hold for the alternative identified as best by the model, so that it occupies one of the top two ranks in the true rank order with probability 90%. In particular therefore, a strategy of employing an additive approximation to produce a shortlist of two or three alternatives might be very useful.

The two models attempting to reduce risk to an expected attribute value perform very differently. If sufficient care is taken to approximate the expected attribute values closely, then the results are good even if some anchoring or other contamination occurs. The expected Z model identifies the true best alternative as either first or second in 70% of the simulations. However, if a lack of rigour and effort allow for a large degree of contamination, then the results are generally poorer than any of the other models. In particular the model ignoring risk places the true best alternative in the top three ranks in less than 50% of the simulations. This highlights the need for caution whenever risk is reduced to a single measure, but does not rule out or even discourage the use of such simplifying heuristics. On the contrary, it appears as if a diligent estimate of the expected value may produce excellent results. As mentioned before, however, if this can only be done by making explicit all possible future performances, then the additive model might as well be used since it remains comfortably superior.

The scenario-based models perform only averagely in terms of the relative quality of the approximations to the rank orders. Average POSN scores range between 3.5 and

4.4, while the average RANK scores are marginally better and are between 3.3 and 4. The relatively poor level of performance is not overly surprising if one considers that the scenario-based models ignore between 80% and 95% of the attribute information. However, there does appear a real danger that applying a scenario model will provide the DM with an inferior alternative as well as a false sense of security at having incorporated uncertainty into the analysis. The true best and modelled best alternatives appear in the top three of each other's rank orders between 55% and 70% of the time, depending on the scenario selection strategy. The true and modelled best alternatives coincide in only 30% of the simulations. Despite the relatively efficient use of information, these results do not therefore support the general use of scenarios as a viable alternative to a full model, additive or otherwise.

Further, the probabilistic selection policies consistently and comprehensively outperform the filtered selection policy, which in fact performs slightly worse than even a random selection. The policy selecting the most likely scenarios performs comfortably better than the other selection policies, and the rank and probabilistic information indicate that such a model may be plausible when accuracy is not a crucial consideration. The probabilistic selection model is consistently but only marginally better than random scenario selection. As a result, it appears necessary not only to select the most likely scenarios, but to do so with a fair degree of accuracy. In practice one might hope to achieve a degree of accuracy somewhere between that of the most likely and probabilistic selection policies. The poor performance of the filtered selection model may in part be due to the imposed constraint that the status quo scenario be selected first. Nevertheless, the results provide at least an initial warning that a selection policy based on the ideas of diversity advocated in scenario planning might not be a panacea.

It should be emphasised that these observations about the scenario models are drawn solely from the numerical results indicating the quality of approximation of the idealised rank order. We do not take into account other advantages which may accrue to the scenario models, for example those cause-and-effect relationships identified in the construction of the scenarios, which may be of considerable benefit as additional learning tools. The results are therefore not presented as conclusive evidence of any universal inferiority of scenario models. In fact, in much the same way as different MCDM methodologies emphasise different aspects of the decision problem, so do the scenario-based approaches. However, the results are meant to warn that there *are* consequences for ignoring part of the decision problem, and that these consequences are often severe. An explicit warning such as this one is often missing from applications of scenario-based models. In many conventional choice problems, a model selecting an alternative which is only fourth or fifth best out of fifteen is simply not good enough. In all cases the final decision as to whether the resulting loss of accuracy is sufficient to offset the time and structural advantages can only be taken after

considering the context of the problem and the needs of the DM.

Shape of Utility Function

The shape of the utility function is determined by the reference level, τ_j , and the value of the utility function at the reference level, λ_j . Tables 5.3 and 5.4 show the effect of τ_j and λ_j respectively, while table 5.5 shows the increase in POSN and RANK scores when λ_j moves from the low interval to the high interval under conditions of low or high reference levels.

The position of the reference level exerts a tremendous influence over the results. The quality of the rank approximation is considerably worse when reference levels are low i.e. the DM views most of the attribute domain as a gain. This deterioration is true in all models, but is particularly severe in the scenario-based models and the model ignoring risk, for which both the POSN and RANK scores increase by between 1.5 and 2 positions when reference levels are low. This deterioration is particularly important: the performance of the scenario-based models when reference levels are high is mediocre but adequate for certain applications, with POSN and RANK scores generally around 3. In such circumstances the true best alternative is located among the top two model ranks in approximately 60% of the simulations. However, when reference levels are low, the average POSN and RANK scores are closer to 5 out of 15, which would almost certainly be deemed unacceptably poor. In fact the extent of the deterioration is such that the true best alternative is located outside of the top five model ranks in at least 30% of the simulations. The use of the scenario models can therefore at this stage only be advocated if reference levels are high. In contrast, the additive model is the least affected and shows a large degree of robustness, although even then the POSN and RANK scores increase by nearly 0.5 when reference levels decrease from high to low. Nevertheless, even where reference levels are low the top two model ranks contain the true best alternative with a probability of 85%.

The reason for these results appears to be due to the limited compensation implied by a high reference level, which is sometimes extreme due to the model requirement that the slope of the utility function be greater below the reference level. This (often severely) limited compensation has the effect of more clearly separating out the performances of alternatives, so that mediocre alternatives are firmly excluded from consideration. This is particularly important for the scenario approaches, which rely on limited information.

The value of the utility function at the reference level exerts a far lesser influence, and although there is a definitive relationship, the effects on the rank orders are marginal. The average POSN and RANK scores are both lower when λ_j is low relative to when it is high. The order of magnitude of this effect does not exceed 0.3, and is again

least evident in the additive model, for which the scores differ by less than 0.1. The general influence of the λ_j is therefore unlikely to be of significant interest.

There is an interesting but modest interaction between the τ_j and λ_j . Essentially, increases in the λ_j result in the increased scores mentioned above only when reference levels are high. When reference levels are low, any increases are of a lower order of magnitude if observed at all. This interaction is consistent over all models, although it is more evident in the POSN scores than in the RANK scores. As mentioned, the interaction is modest in terms of the difference in rank orders. For the scenario models in which the effect is most visible, increases in the λ_j result in increases of 0.1 in POSN scores when reference levels are low, compared to increases of 0.4 when reference levels are high.

	POSN		RANK		
	τ	low	high	low	high
Additive	1.60	1.19	1.64	1.19	
Ignore	5.73	4.32	5.00	3.66	
Expected Z	2.99	1.92	2.67	1.75	
Scenario R	5.16	3.17	4.95	2.98	
Scenario P	4.91	3.09	4.69	2.93	
Scenario ML	4.14	2.77	4.04	2.62	
Scenario F	5.40	3.32	4.98	3.08	

Table 5.3: Effect of reference level (τ_j)

	POSN		RANK		
	λ	low	high	low	high
Additive	1.36	1.44	1.37	1.45	
Ignore	4.93	5.11	4.24	4.42	
Expected Z	2.41	2.50	2.11	2.31	
Scenario R	4.05	4.28	3.83	4.10	
Scenario P	3.88	4.12	3.67	3.95	
Scenario ML	3.37	3.54	3.20	3.45	
Scenario F	4.25	4.47	3.90	4.16	

Table 5.4: Effect of utility at reference level (λ_j)

		POSN	RANK
Additive	low τ_j	+0.03	+0.04
	high τ_j	+0.12	+0.12
Ignore	low τ_j	+0.06	+0.06
	high τ_j	+0.30	+0.30
Expected Z	low τ_j	-0.02	+0.21
	high τ_j	+0.21	+0.19
Scenario R	low τ_j	+0.04	+0.09
	high τ_j	+0.41	+0.44
Scenario P	low τ_j	+0.09	+0.15
	high τ_j	+0.41	+0.42
Scenario ML	low τ_j	+0.10	+0.18
	high τ_j	+0.25	+0.32
Scenario F	low τ_j	+0.06	+0.13
	high τ_j	+0.38	+0.40

Table 5.5: Effect of interaction between reference level (τ_j) and utility at the reference level (λ_j)

Attribute Variability

The POSN and RANK scores for each simulated degree of attribute variability are shown in table 5.6, while table 5.7 shows the relative increases in POSN and RANK scores when attribute variability increases under conditions of low or high reference levels. The variability of the attribute values also exerts a strong influence over the results. Increased attribute variability results in some large deteriorations in the quality of the rank approximations. Excluding the additive model, which is not influenced by the different levels of variability, the magnitude of the deteriorations ranges from 0.5 for the expected Z model to over 1 position for the filtered scenario model. The deteriorations are fairly consistent over both the POSN and RANK scores, and are large enough to warrant further attention. The deterioration manifests itself in the rank probability matrix by decreasing the probability that the true and modelled best alternatives coincide by around 10% when attribute variability increases. Such deteriorations are relatively more harmful to the scenario-based methods, where the probability that the best alternative is agreed upon is only 35% when attribute variability is low. By way of comparison the same event has a probability of 55% when the expected Z model is used.

It seems fairly clear that the level of variability would have a significant effect for those methods which consider only a fraction of the total information: the smaller the variability, the smaller the sample size required to capture the essence of the total population. A less obvious and possibly more interesting observation is derived from the poor performance of the filtered scenario model when attribute variability

is high. It is often argued from a scenario planning standpoint that, when future outcomes are highly divergent, the scenario planning techniques become superior to the probabilistic methods (for example, [101]). Yet the results show that the relative deteriorations due to increased attribute variability are most keenly felt in the filtered scenario model. This is another early indication that uncritical application of scenario planning techniques to MCDM might be misguided and detrimental to results.

There is a small but fairly consistent interaction between the effects of attribute variability and reference levels in the POSN scores. It appears that the deterioration in results due to increased attribute variability is more pronounced when reference levels are high than when they are low. This observation is only true for the model ignoring risk and the scenario models; table 5.7 therefore excludes the additive and expected Z models. Again the magnitude of the interaction effect is marginal; deteriorations are between 0.3 and 0.4 positions more severe when reference levels are high.

B	POSN		RANK	
	low	high	low	high
Additive	1.38	1.42	1.39	1.44
Ignore	4.65	5.40	3.92	4.74
Expected Z	2.18	2.73	2.00	2.42
Scenario R	3.67	4.66	3.49	4.43
Scenario P	3.55	4.46	3.35	4.27
Scenario ML	3.12	3.79	2.98	3.68
Scenario F	3.82	4.91	3.52	4.55

Table 5.6: Effect of attribute variability (B)

		POSN	RANK
Ignore	low τ_j	+0.45	+0.80
	high τ_j	+1.05	+0.84
Scenario R	low τ_j	+0.88	+0.96
	high τ_j	+1.08	+0.92
Scenario P	low τ_j	+0.76	+0.91
	high τ_j	+1.06	+0.95
Scenario ML	low τ_j	+0.47	+0.61
	high τ_j	+0.88	+0.79
Scenario F	low τ_j	+0.95	+0.99
	high τ_j	+1.22	+1.06

Table 5.7: Effect of interaction between reference level (τ_j) and attribute variability (B)

Multivariate Risk Proneness

The degree of multivariate risk proneness is determined by the parameter κ . Table 5.8 shows the effect of various values of κ on the performance measures. The level of multivariate risk proneness has only a moderate effect on results for even quite large degrees of proneness. The quality of the average POSN scores do deteriorate as the multivariate risk proneness coefficient κ increases in magnitude, but the effects of such deteriorations are not materially experienced until $\kappa = 8$. Even then, the deteriorations relative to the case when $\kappa = 1$ are only of order 0.3, and so can be considered modest. Deteriorations in RANK scores are only exhibited by the additive model. In contrast to the previously mentioned effects, the additive model is subject to the most severe deteriorations due to increases in multivariate risk proneness. The relative frequency with which the true and modelled best alternatives coincide in the additive model decreases from 85% when $\kappa = 1$ to 65% when $\kappa = 8$. However, the additive approximation remains excellent, the true best alternative still being ranked either first or second with probability greater than 80%, and the modelled best alternative also being either first or second in the true rank order with similar probability. This is true even for large κ , which supports similar results reported by Stewart [85].

There is a strong interaction between the effects of reference levels and κ for the additive model only. Table 5.9 therefore shows the POSN and RANK scores resulting from the interaction between multivariate risk proneness and reference levels for the additive model only. The deterioration due to increasing multivariate risk proneness is far greater when reference levels are low. In fact, increasing multivariate risk proneness has little or no impact when reference levels are high, resulting in deteriorations in both the POSN and RANK scores of order 0.1 when κ increases from 1 to 8. However, those same increases result in deteriorations of order 0.8 when reference levels are low. As a result, some fairly average results are observed when $\kappa = 8$ and reference levels are low, with the true best alternative being first in the modelled rank order in only 50% of the simulations and outside the top two ranks in 30%. There is no obvious explanation for the behaviour resulting from this interaction. We may observe that when reference levels are low, the greater part of the partial utility functions are concave i.e. risk averse. Without confusing the notions of risk aversion in the monocriterion and multivariate senses, it is possible that increasing multivariate risk proneness conflicts with the criterion-wise risk aversion, resulting in the deteriorations that are experienced. The general nature of this interaction is evident in the POSN scores of the other models, albeit to a far lesser degree. The magnitude of the interactions in these other models is not considered sufficient to warrant deeper investigation.

	κ	POSN			RANK		
		1	3	8	1	3	8
Additive		1.21	1.31	1.68	1.22	1.32	1.70
Ignore		4.98	4.98	5.11	4.45	4.31	4.23
Expected Z		2.34	2.38	2.64	2.18	2.16	2.29
Scenario R		4.07	4.13	4.30	4.01	3.95	3.93
Scenario P		3.91	3.93	4.17	3.84	3.76	3.83
Scenario ML		3.37	3.40	3.60	3.34	3.31	3.33
Scenario F		4.23	4.33	4.53	4.06	4.01	4.02

Table 5.8: Effect of multivariate risk proneness (κ)

	κ	POSN			RANK		
		1	3	8	1	3	8
Additive	low τ_j	1.29	1.46	2.06	1.30	1.49	2.12
	high τ_j	1.13	1.15	1.30	1.13	1.15	1.29

Table 5.9: Effect of interaction between reference level (τ_j) and multivariate risk proneness (κ)

Number of Scenarios

The number of scenarios chosen, denoted by s , will obviously affect only those models which are dependent on the scenario selection policies. As a result we exclude the other three models from consideration in table 5.10, which gives the POSN and RANK scores for each value of s . For the scenario models, the number of scenarios selected exercises a considerable influence over results. Increasing the number of scenarios included in the analysis from three to five results in improvements in the scores of both rank orders of order 0.5. Increasing the number to ten scenarios results in further improvements of order 0.9. Generally though, the performance of the scenario-based models remains poor. If 5 out of the 50 possible scenarios are considered, the true best alternative is on average ranked fourth or worse in all scenario models except that which considers the most likely scenarios. The performance is considerably more acceptable in the scenario model selecting the most likely scenarios, with POSN and RANK scores as low as 2.5 when 10 scenarios are considered. In such circumstances the true best alternative is located either first or second in the modelled rank order in nearly 70% of simulations. Even in the other selection models, the same event occurs with a probability approaching 60%. Although it is doubtful whether adequate attention can be given to 20% of the scenarios, the results do give at least some numerical justification for using some kind of scenario-based approach where the aim is the production of a shortlist of 2 or 3 alternatives. Nevertheless, it should be reiterated that the performance of the scenario models is in most circumstances poor.

An interesting feature is that although the initial improvements due to increasing the number of scenarios from three to five is quite different across the selection policies, the further gains obtained by increasing the number of scenarios to ten is almost exactly the same in all selection policies. The rates of marginal improvement are thus greatest in those selection policies relying on the probabilities of occurrence. In particular it appears that in such models the improved results justify the effort involved in constructing five scenarios instead of three.

<i>s</i>	POSN			RANK		
	3	5	10	3	5	10
Scenario R	4.84	4.26	3.40	4.58	4.05	3.26
Scenario P	4.73	4.07	3.20	4.47	3.86	3.10
Scenario ML	4.22	3.51	2.64	4.02	3.40	2.56
Scenario F	4.92	4.52	3.64	4.36	4.25	3.50

Table 5.10: Effect of the number of selected scenarios (*s*)

General Robustness of Results

Having examined the significant and important effects, we may make some statements about the general robustness of the models towards the parameters of this simulation experiment. As a simple vehicle for these statements we make use of the rank probability matrix to construct the probability that the true best alternative is located either first or second in the modelled rank order. It is then possible to see to what extent results may deteriorate or improve if the effects previously discussed conspire to form 'worst' and 'best' cases. For example, the previous discussions indicated that the worst results i.e. the aforementioned 'worst' case, would occur if the set of parameters combined so that reference levels were low, utilities at the reference levels were high, attribute variability was high, multivariate risk proneness was high and the number of scenarios chosen was low. Similarly, a 'best' case can be found by considering the performance measures when, simultaneously, reference levels are high, utilities at the reference levels are low, attribute variability is low, multivariate risk proneness is low and the number of scenarios chosen is high. An 'average' performance is constructed as before by averaging over all cases. Table 5.11 shows the resulting worst, average, and best possible probability that the true best alternative is among the top two alternatives in each of the model rank orders. The non-scenario models are by definition independent of the number of scenarios chosen, while the scenario models themselves are obviously non-robust to such changes. In order to make meaningful comments about the relative robustness of the models it is therefore necessary to avoid penalising the scenario models on the basis of a parameter that must be fixed at the outset of the analysis. Furthermore, the issue of comparing robustness is complicated by the different levels of general performance. In such circumstances, it is not clear to what extent the apparent insensitivity of a weaker model is due to actual notions of robustness or just general poor performance. Therefore, although

it implies ignoring all interactions involving the number of scenarios, the number of selected scenarios is held constant at the maximum of 10 in order firstly to avoid understating the robustness of the scenario models and, secondly, to minimise any effects of poor performance that might cloud the robustness relationships.

	Worst	Average	Best
Additive	0.72	0.91	0.99
Ignore	0.27	0.37	0.51
Expected Z	0.50	0.69	0.86
Scenario R	0.33	0.55	0.80
Scenario P	0.38	0.58	0.81
Scenario ML	0.48	0.65	0.83
Scenario F	0.31	0.52	0.78

Table 5.11: Probabilities that the true best alternative is placed first or second in the model rank order

It is generally the best performing models that are also the most robust to changes in the parameters. In particular, the additive model appears to be comfortably superior to all the other models under any circumstances. The expected Z and scenario ML models are subject to slightly larger deteriorations than the additive model as the parameters take on less favourable values, but remain fairly robust, performing adequately until conditions become fairly extreme. The most striking aspect of the results is the poor robustness of the other scenario models. In fact, all the scenario models perform well when conditions are most favourable, with results only marginally worse than the expected Z model. Moreover, the relative performances of the four selection policies are similar. However, as the parameters take on less favourable values the performances of the three scenario models deteriorate far quicker than the other models, probably due to their sensitivity to reference levels. Furthermore, as conditions deteriorate the superiority of probabilistic scenario selection, and in particular the most likely selection policy, becomes increasingly evident. As a result we observe that the robustness of the scenario models increases with an increasing reliance on probability as a selection criterion.

Finally, it is evident that the model ignoring risk is subject to the least deterioration, at least from an absolute point of view. However, a label of robustness is hesitantly used, as it is not clear to what extent the apparent robustness is a product of the general poor performance of this model. It certainly appears to be genuinely more robust than the scenario approaches, as the relative superiority of the scenario models decreases markedly as conditions move away from the ideal. Comparisons to the additive and expected Z models with regard to robustness appear less important given the magnitude of superiority enjoyed by those models under all conditions.

5.1.3 Summarised Inferences

In conclusion, the following messages are observed in the results and formalised as summarised inferences:

- The quality of approximation is considerably worse when preferences are mainly compensatory i.e. reference levels are low. The result is particularly strong in the scenario models, which exhibit very poor performance when reference levels are low.
- A slightly less powerful but still important effect is related to the variability of the attribute values over scenarios. Results are materially degraded when attribute variability is high in all models except the additive model.
- Contrary to conventional scenario planning wisdom, when attribute values become more variable the selection of divergent scenarios results in a relatively greater degradation of results than if scenarios are selected probabilistically.
- Results are unlikely to be substantially affected by multivariate risk proneness unless such proneness is extreme and reference levels are low.
- There are quite substantial marginal improvements resulting from considering five rather than three scenarios. The marginal improvements are greatest where the most likely scenarios are chosen, and become less attractive as the probability of occurrence is replaced by diversity as the basis for scenario selection.
- The most successful models, particularly the additive model, are also the most robust to changes in the parameters of experiment 1. The scenario models, with the exception of the most likely selection policy, are particularly sensitive to changes in the parameters, performing well under favourable conditions but poorly or very poorly under others.
- Robustness in the scenario models increases with dependency on the probability of occurrence as a basis for scenario selection.

Effect	DoF	Additive	Ignore	Expected Z	Scenario R	Scenario P	Scenario ML	Scenario F
τ_j	1	3619.70**	2690.20**	5526.64**	7570.47**	5527.69**	4302.61**	7137.69**
λ_j	1	125.37**	44.03**	43.82**	99.14**	101.99**	70.73**	77.63**
B	1	35.82**	767.03**	1468.68**	1833.27**	1381.88**	1035.10**	1944.23**
s	2	0.31	0.06	0.94	1331.34**	1319.02**	1898.58**	938.22**
k	2	1745.81**	10.46**	171.24**	35.00**	45.79**	46.36**	51.87**
$\tau_j \times \lambda_j$	1	38.90**	18.61**	59.72**	64.17**	42.14**	12.39**	41.52**
$\tau_j \times B$	1	28.80**	120.93**	4.81*	19.56**	35.43**	93.96**	30.12**
$\lambda_j \times B$	1	0.29	5.29*	6.71*	0.53	0.36	2.23	0.00
$\tau_j \times s$	2	0.36	0.85	1.31	6.84**	4.74*	21.24**	1.04
$\lambda_j \times s$	2	1.31	2.24	1.51	2.73	0.56	0.49	0.38
$B \times s$	2	0.35	1.85	1.99	2.47	2.63	19.75**	5.57**
$\tau_j \times k$	2	704.46**	0.76	1.57	9.02**	12.97**	12.60**	33.97**
$\lambda_j \times k$	2	5.24*	0.77	1.99	0.69	0.04	0.33	0.13
$B \times k$	2	28.05**	0.70	8.11**	0.68	0.21	0.60	0.48
$s \times k$	4	0.15	0.12	0.66	1.56	1.46	2.11	0.86
$\tau_j \times \lambda_j \times B$	1	25.58**	1.98	76.79**	7.64*	22.58**	5.39*	33.58**
$\tau_j \times \lambda_j \times s$	2	1.54	1.71	2.75	0.13	3.06*	0.24	0.12
$\tau_j \times B \times s$	2	0.03	0.96	1.90	14.04**	21.38**	6.56**	12.89**
$\lambda_j \times B \times s$	2	0.07	0.22	0.22	0.46	2.60	1.43	0.63
$\tau_j \times \lambda_j \times k$	2	92.79**	4.47*	34.53**	6.81**	3.75*	0.74	0.28
$\tau_j \times B \times k$	2	4.99*	0.11	13.15**	5.57**	2.56	0.62	0.27
$\lambda_j \times B \times k$	2	0.07	0.33	0.66	0.93	0.33	0.56	2.61
$\tau_j \times s \times k$	4	0.05	0.62	1.12	0.95	0.57	1.72	3.00*
$\lambda_j \times s \times k$	4	2.44*	2.16	2.39	0.63	1.55	1.20	0.93
$B \times s \times k$	4	0.05	1.21	1.51	0.60	2.18	2.16	0.91
$\tau_j \times \lambda_j \times B \times s$	2	0.65	0.68	2.13	2.47	0.08	2.16	0.36
$\tau_j \times \lambda_j \times B \times k$	2	14.33**	0.25	8.60**	1.43	0.80	0.43	0.36
$\tau_j \times \lambda_j \times s \times k$	4	0.99	0.41	1.46	2.02	0.26	1.19	1.88
$\tau_j \times B \times s \times k$	4	0.32	1.07	0.92	0.75	0.31	0.65	0.81
$\lambda_j \times B \times s \times k$	4	0.93	1.64	1.41	1.38	0.63	0.05	1.11
$\tau_j \times \lambda_j \times B \times s \times k$	4	0.06	2.18	1.75	1.07	0.08	1.47	0.46

Table 5.12: ANOVA results for POSN scores in experiment 1

Effect	DoF	Additive	Ignore	Expected Z	Scenario R	Scenario P	Scenario ML	Scenario F
τ_j	1	3804.15**	3272.04**	5408.81**	7340.84**	5444.96**	5328.90**	6958.18**
λ_j	1	115.73**	61.40**	249.79**	135.14**	143.15**	162.58**	132.82**
B	1	39.81**	1234.40**	1123.92**	1657.13**	1504.09**	1295.06**	2028.15**
s	2	1.12	0.07	3.39*	1092.38**	1106.20**	1899.25**	563.33**
k	2	1664.75**	30.20**	38.20**	5.05*	4.47*	0.72	2.08
$\tau_j \times \lambda_j$	1	27.96**	27.44**	0.39	56.80**	33.28**	12.36**	34.96**
$\tau_j \times B$	1	16.97**	0.62	6.50*	0.61	0.65	19.63**	2.79
$\lambda_j \times B$	1	1.93	15.02**	0.06	7.26*	1.74	3.99*	22.78**
$\tau_j \times s$	2	0.45	1.60	0.28	11.55**	11.51**	54.37**	24.18**
$\lambda_j \times s$	2	0.23	0.83	0.91	6.25**	2.13	2.59	2.90
$B \times s$	2	1.39	0.14	1.41	1.75	8.20**	21.53**	1.86
$\tau_j \times k$	2	750.84**	12.67**	19.89**	0.44	2.43	0.17	4.90*
$\lambda_j \times k$	2	5.49**	0.82	4.46*	0.91	0.23	0.77	0.56
$B \times k$	2	30.33**	1.60	0.94	1.37	0.62	0.89	0.98
$s \times k$	4	0.33	0.70	1.85	3.25*	1.76	3.10*	1.72
$\tau_j \times \lambda_j \times B$	1	29.10**	0.93	10.76**	6.94*	10.46**	6.57*	6.19*
$\tau_j \times \lambda_j \times s$	2	0.82	0.93	0.55	1.35	4.85*	0.85	2.69
$\tau_j \times B \times s$	2	0.51	0.60	1.58	2.43	4.18*	5.04*	7.26**
$\lambda_j \times B \times s$	2	0.68	0.07	0.02	0.34	0.41	1.07	1.98
$\tau_j \times \lambda_j \times k$	2	71.82**	1.57	25.70**	0.22	0.98	0.22	0.11
$\tau_j \times B \times k$	2	3.65*	1.22	11.83**	0.21	0.91	0.83	0.08
$\lambda_j \times B \times k$	2	1.83	0.93	0.69	0.14	1.12	1.13	2.15
$\tau_j \times s \times k$	4	0.30	0.95	0.33	1.48	0.74	2.43*	0.23
$\lambda_j \times s \times k$	4	0.47	1.01	1.23	0.81	0.71	0.80	0.38
$B \times s \times k$	4	0.62	0.69	0.37	0.59	0.24	0.73	0.22
$\tau_j \times \lambda_j \times B \times s$	2	0.65	0.85	0.78	0.63	0.05	3.22*	2.47
$\tau_j \times \lambda_j \times B \times k$	2	18.18**	5.31*	0.30	0.44	0.55	0.39	1.55
$\tau_j \times \lambda_j \times s \times k$	4	0.17	0.45	1.31	0.78	0.30	0.48	0.71
$\tau_j \times B \times s \times k$	4	0.28	2.17	0.18	0.62	1.52	0.27	0.30
$\lambda_j \times B \times s \times k$	4	1.41	2.60*	1.56	0.28	0.12	0.24	0.41
$\tau_j \times \lambda_j \times B \times s \times k$	4	0.92	0.89	0.93	1.89	0.46	0.57	0.11

Table 5.13: ANOVA results for RANK scores in experiment 1

5.2 Results of Simulation Experiment 2

Simulation experiment 2 investigates the impact of two non-idealities relating to elicitation or assessment: errors in the assessment of the criterion weights, and errors in the assessment of the scenario probabilities. There are therefore four main effects, one corresponding to the variability of the attribute values (B), one corresponding to the number of scenarios chosen (s), and two corresponding to the assessment errors (e_w and e_{Pr}). The shape of the utility function is fixed at case 1 in order to limit the scope of the investigation to the 54 cases corresponding to combinations of the following parameters

Parameter	Values
B	low, high
e_w	0, U(-0.05, 0.05), U(-0.1, 0.1)
e_{Pr}	0, U(-0.3, 0.3), U(-0.6, 0.6)
s	3, 5, 10

Table 5.14: Parameters for experiment 2

5.2.1 ANOVA Results

The detailed ANOVA tables for experiment 2 can be found at the end of section 5.2 in tables 5.22 and 5.23, for the POSN and RANK scores respectively. There is again a high degree of consistency between the models in terms of significant main and higher-order effects. The most significant effects are those relating to the variability of the attribute values and the number of scenarios selected. As in experiment 1, the F-statistics of these effects are of similar magnitude. However, as they have already been discussed, their role in this experiment is limited to any interactions they may be involved in with the assessment errors. Of the assessment errors, those relating to criterion weights have the most significant effect. The F-statistics vary quite considerably between models, being most significant in the additive and expected Z model, but are significant at the 0.5% level in all the models. The effect of probability assessment errors is less significant, with F-statistics generally around 10, and there is some inconsistency between the POSN and RANK results, with the errors generally being less significant in the RANK results.

There are two clear systematic two-way interaction effects: between attribute variability and weight assessment errors, and between the number of scenarios and weight assessment errors. The former is significant in all models, while the latter, involving the number of scenarios, is significant only in the probabilistic scenario-based models. The interaction is not significant for the filtered scenario selection policy. Further, there are no higher-order interactions.

5.2.2 Characterisation of Effects

An overview of the effects of assessment errors can be obtained by comparing the grand average scores in simulation 2 with results when no assessment errors are made. To this end we make use of the results of simulation 1, in which all assessment errors were fixed at zero. In order to ensure that the displayed results differ only with respect to the presence or absence of assessment errors, it is necessary to consider only certain cases in the first simulation i.e. those for which τ_j and λ_j are low and $\kappa = 3$. The following table shows the POSN and RANK scores in both circumstances, together with the rank correlations.

Model	POSN _{no error}	POSN _{error}	RANK _{no error}	RANK _{error}
Additive	1.42	2.07	1.45	2.13
Ignore	5.69	5.75	4.96	5.16
Expected Z	2.91	3.24	2.49	2.93
Scenario R	5.10	5.25	4.86	5.04
Scenario P	4.73	4.97	4.51	4.77
Scenario ML	4.04	4.27	3.94	4.24
Scenario F	5.32	5.45	4.87	5.05

Table 5.15: Average POSN and RANK scores for experiment 2 (error) and corresponding experiment 1 (no error) problem contexts

Model	SRC _{no error}	SRC _{error}
Additive	0.95	0.83
Ignore	0.30	0.29
Expected Z	0.70	0.65
Scenario R	0.40	0.37
Scenario P	0.44	0.41
Scenario ML	0.52	0.48
Scenario F	0.37	0.35

Table 5.16: Average rank correlations for experiment 2 (error) and corresponding experiment 1 (no error) problem contexts

The magnitude of the general deteriorations are not particularly pronounced relative to the effects already discussed, although they clearly need to be further categorised by the size of the actual assessment errors to gain an insight into how large the potential deteriorations might be. The deterioration in the results of the additive model is the most marked; for all other models the deteriorations are fairly limited. The deteriorations in the rank correlations are particularly small. This is an early indication that errors in the assessments of criterion weights and scenario probabilities may not be crucial to the accuracy of results.

Errors in Criterion Weight Assessment

The POSN and RANK scores resulting from various degrees of error in assessing criterion weights are shown in table 5.17. Further, two interaction effects involving weight assessment errors are discussed: table 5.18 shows the deterioration of results when assessment errors are present relative to the case where none are made, under high and low attribute variability, and table 5.19 shows the deterioration of results due to weight assessment errors relative to the case where no errors are made, when different numbers of scenarios are selected.

The weight assessment errors exert quite different degrees of influence over the various models. In general, the weight assessment errors cause marginally greater increases in the RANK scores than the POSN scores. The additive model experiences notable increases of order 0.4 in both POSN and RANK scores for even moderate errors, while scores are further downgraded by 0.85 when assessment errors are more substantial. While for moderate assessment errors the probability that the additive model ranks the true best alternative first or second remains over 75%, the figure drops to 60% when greater errors are made. Under such circumstances, the third-ranked alternative would need to be included in the shortlist in order to push the probability of including the true best alternative to 75%.

On the other hand, deteriorations in the quality of results in the scenario-based models and the model ignoring risk are very limited for moderate assessment errors. For assessment errors of the order 0.05, the scores increased by less than 0.1. Small errors in the weight assessment process therefore seem particularly inconsequential in a scenario-based environment. Even for more substantial errors of order 0.1, the resultant deterioration was marginal at between 0.2 and 0.3. Regardless of weighting errors, the relative frequency with which the true and modelled best alternatives coincide hovers around only 25%, while the true best alternative is ranked outside the top five model ranks in between 30% and 40% of the simulations. The probability-based selection models were slightly more sensitive to the weighting errors than were either the random or filtered models. The influence of weighting errors on the expected Z model is somewhere between that of the additive model and the model ignoring risk. For smaller errors, deteriorations are of order 0.25 and may be considered modest. However, when assessment errors are more pronounced, some fairly substantial deteriorations (0.65 for POSN, 0.85 for RANK) can arise. Under such conditions the performance of the expected Z model becomes relatively closer to that of the additive model, and the probability of identifying the true best alternative among the top two model ranks is only 10% lower than the additive model at just over 50%. What is therefore evident is that it is the more accurate models that are more sensitive to weight assessment errors.

An interaction between the effects of attribute variability and the errors in assessing the weights is observed as a consistent but small-scale result: errors in the assessment of weights are felt more strongly when attribute variability is low. Although in most instances the magnitude of these differences might not be of practical importance, they are consistently observed in all the scenario models and the expected Z model. In these models, the deterioration of scores due to weight assessment errors is between 0.3 and 0.4 positions worse when attribute variability is low. The other two models, the additive model and the model ignoring risk, experienced no such interaction effect, and are consequently excluded from table 5.18.

The other observed interaction occurs between the effects of the number of selected scenarios and the errors in assessing the weights, although the consistent progression of the results makes for more convincing interpretation: the effects of errors in the assessment of weights are felt more strongly when a larger number of scenarios are selected for analysis, regardless of which selection strategy is employed. The differences due to the interaction effect are of order 0.4 for the probabilistic selection policies and 0.2 for the filtered selection policy, and so cannot be expected to play a crucial role in the quality of results.

	e_w	POSN			RANK		
		0	0.05	0.1	0	0.05	0.1
Additive	1.53	1.91	2.76	1.53	1.97	2.90	
Ignore	5.67	5.72	5.86	5.02	5.12	5.33	
Expected Z	2.97	3.15	3.61	2.54	2.83	3.41	
Scenario R	5.13	5.16	5.45	4.92	4.98	5.23	
Scenario P	4.84	4.92	5.15	4.64	4.66	4.99	
Scenario ML	4.14	4.21	4.45	4.10	4.15	4.46	
Scenario F	5.35	5.43	5.56	4.87	5.00	5.29	

Table 5.17: Effect of errors in the assessment of criterion weights (e_w)

		POSN		RANK		
		e_w	0.05	0.1	0.05	0.1
Expected Z	low B		+0.22	+0.56	+0.35	+0.69
	high B		+0.14	+0.35	+0.24	+0.45
Scenario R	low B		+0.07	+0.44	+0.15	+0.34
	high B		+0.01	+0.13	-0.03	+0.17
Scenario P	low B		+0.09	+0.33	+0.01	+0.46
	high B		+0.08	+0.11	+0.03	+0.20
Scenario ML	low B		+0.14	+0.29	+0.12	+0.36
	high B		+0.00	+0.19	-0.02	+0.25
Scenario F	low B		+0.23	+0.14	+0.21	+0.41
	high B		-0.08	+0.14	+0.05	+0.17

Table 5.18: Effect of interaction between attribute variability (B) and weight assessment errors (e_w)

		POSN		RANK		
		e_w	0.05	0.1	0.05	0.1
Scenario R	3		-0.07	+0.17	-0.03	+0.18
	5		+0.05	+0.23	+0.03	+0.23
	10		+0.14	+0.47	+0.17	+0.36
Scenario P	3		+0.03	+0.13	-0.05	+0.24
	5		+0.04	+0.21	-0.03	+0.33
	10		+0.19	+0.32	+0.13	+0.41
Scenario ML	3		-0.03	+0.14	-0.05	+0.22
	5		+0.14	+0.20	+0.01	+0.28
	10		+0.10	+0.38	+0.19	+0.42
Scenario F	3		+0.14	+0.01	+0.11	+0.27
	5		+0.00	+0.14	+0.08	+0.25
	10		+0.08	+0.27	+0.20	+0.34

Table 5.19: Effect of interaction between the number of selected scenarios (s) and weight assessment errors (e_w)

Errors in Scenario Probability Assessment

In contrast to the weight assessment errors, errors in the assessment of the scenario probabilities exercise only a trivial influence. Table 5.20 shows the POSN and RANK scores under various degrees of probability assessment error. The largest errors of 0.6 in the assessment of probabilities resulted in deteriorations in scores of not more than 0.25, and were in most cases closer to 0.1. Smaller errors had almost no effect at all. The most sensitive models are the additive model, the expected Z model and the scenario model selecting the p most likely scenarios. The sensitivity of the first

two models has been explained previously as a type of diminishing marginal return, so that the best performing models suffer more from non-idealities. The deterioration of the results using the most likely selection policy also might be expected given the extreme dependence of this method on the probabilities of occurrence. Nevertheless, it still comfortably outperforms all other scenario models. In any case the marginal deteriorations in the models are unlikely to be sufficiently large to be of practical importance.

	POSN			RANK			
	e_{Pr}	0	0.3	0.6	0	0.3	0.6
Additive		1.99	2.04	2.19	2.05	2.11	2.23
Ignore		5.71	5.78	5.76	5.13	5.16	5.18
Expected Z		3.16	3.25	3.32	2.85	2.92	3.01
Scenario R		5.22	5.21	5.31	5.03	5.05	5.05
Scenario P		4.91	4.93	5.06	4.76	4.73	4.80
Scenario ML		4.14	4.30	4.37	4.14	4.23	4.34
Scenario F		5.41	5.44	5.49	5.01	5.05	5.10

Table 5.20: Effect of errors in the assessment of scenario probabilities (e_{Pr})

General Robustness of Results

Again we conclude the discussion of this simulation experiment by examining the relative robustness of each model. The discussions in the section thus far have identified that the quality of the rank approximation deteriorates when attribute variability is high, errors in the assessment of criterion weights and scenario probabilities are highest and the number of scenarios included is lowest. Table 5.21 shows the corresponding worst-case probability that the true best alternative is ranked in the first or second model positions, with similar average and best-case probabilities for each model.

	Worst	Average	Best
Additive	0.58	0.76	0.91
Ignore	0.25	0.30	0.34
Expected Z	0.46	0.56	0.67
Scenario R	0.31	0.40	0.49
Scenario P	0.36	0.43	0.53
Scenario ML	0.44	0.52	0.62
Scenario F	0.28	0.37	0.46

Table 5.21: Probabilities that the true best alternative is placed first or second in the model rank order

It appears that the better-performing models are no longer more robust to the changes in parameter values, and all models experience roughly similar relative deteriorations

of 30%. There are some marked changes in the relative order of robustness: the additive model is now one of the least robust models, while the scenario ML model is one of the most robust. In fact, there are two conflicting robustness relationships in operation. From experiment 1 it was possible to conclude that the scenario-based models were more sensitive to changes in attribute variability, which is also simulated in experiment 2. However, the previous discussions relating to this experiment have identified that it is the additive and expected Z models that are more sensitive to the other two assessment error effects. It is this increased sensitivity that results in the relatively less robust behaviour of the additive and expected Z model. In fact, the additive model is particularly adversely affected, so that there is little to choose from between the additive and expected Z models when both assessment errors are high.

The scenario models are far more robust than in experiment 1 due to the exclusion of the reference levels to which they were particularly sensitive, as well as their relative robustness to assessment errors. The increased robustness of poorly-performing models to the assessment errors means that the robustness of the scenario model selecting the most likely scenarios has declined relative to the other models. However, the consistent superiority of the scenario ML model in the context of experiment 2 makes it difficult to explore the robustness relationships more deeply. Even worse is the performance of the model ignoring risk, which although appearing as the most robust of the models performs so badly under even favourable conditions that practical acceptability is unlikely.

5.2.3 Summarised Inferences

The results of simulation experiment are condensed into the following inferences:

- The criterion weight assessment errors must be fairly severe before marked deterioration in the rank orderings occur. Small to moderate assessment errors have only marginal effect on the results.
- The deteriorations caused by errors in the assessment of scenario probabilities are too small to have any practical effect on the quality of the results.
- Deteriorations due to weight assessment errors are more prominent when attribute variability is low.
- In the scenario models, deteriorations due to weight assessment errors are more prominent when a larger set of scenarios is included in the analysis.
- Models which generally perform better, specifically the additive and expected Z model, are more sensitive to both assessment errors than those models that perform poorly.

Effect	DoF	Additive	Ignore	Expected Z	Scenario R	Scenario P	Scenario ML	Scenario F
B	1	0.81	174.59**	792.63**	536.76**	682.25**	280.75**	695.66**
e_w	2	2438.44**	10.39**	286.38**	32.17**	36.69**	44.74**	12.22**
e_{Pr}	2	66.57**	1.47	16.97**	3.33*	9.70**	22.58**	1.82
s	2	1.16	0.54	0.49	519.51**	771.19**	1094.72**	384.54**
$B \times e_w$	2	4.28*	0.02	14.73**	10.05**	6.12**	6.77**	8.88**
$B \times e_{Pr}$	2	7.70**	3.02	1.51	0.43	0.58	0.49	0.92
$e_w \times e_{Pr}$	4	0.83	0.04	0.07	0.72	2.94*	0.81	0.03
$B \times s$	2	0.50	2.69	0.34	2.96	0.28	0.59	1.46
$e_w \times s$	4	1.82	1.37	2.15	5.85**	3.95**	5.56**	2.16
$e_{Pr} \times s$	4	0.65	0.50	0.41	0.62	2.46*	0.38	0.87
$B \times e_w \times e_{Pr}$	4	0.85	0.45	1.14	0.30	0.84	0.17	0.82
$B \times e_w \times s$	4	0.03	0.25	0.69	0.46	1.01	1.37	0.31
$B \times e_{Pr} \times s$	4	0.79	1.14	1.06	2.23	1.01	2.29	1.43
$e_w \times e_{Pr} \times s$	8	0.91	0.97	1.17	1.31	0.92	0.99	1.42
$B \times e_w \times e_{Pr} \times s$	8	1.05	1.34	0.95	1.38	1.18	0.54	0.58

Table 5.22: ANOVA results for POSN scores in experiment 2

Effect	DoF	Additive	Ignore	Expected Z	Scenario R	Scenario P	Scenario ML	Scenario F
B	1	3.25	507.81**	489.19**	890.83**	761.13**	343.57**	952.81**
e_w	2	2575.25**	38.72**	613.70**	40.45**	54.21**	59.32**	66.46**
e_{Pr}	2	41.98**	0.92	19.16**	0.29	1.84	15.88**	3.33*
s	2	1.80	0.47	0.34	660.71**	661.62**	942.71**	256.71**
$B \times e_w$	2	5.70**	4.36*	25.94**	10.66**	7.23**	6.03**	14.82**
$B \times e_{Pr}$	2	2.88	0.16	0.70	0.48	0.73	0.02	0.71
$e_w \times e_{Pr}$	4	0.42	0.44	0.44	0.63	0.37	1.71	1.12
$B \times s$	2	0.40	3.26*	1.41	0.40	0.22	0.25	3.45*
$e_w \times s$	4	0.92	0.43	1.69	4.75**	3.98**	6.56**	1.45
$e_{Pr} \times s$	4	1.69	2.36	1.85	2.40	0.68	2.07	0.48
$B \times e_w \times e_{Pr}$	4	0.31	0.97	0.64	1.31	2.41	1.05	0.46
$B \times e_w \times s$	4	1.48	0.51	1.57	0.85	1.18	2.45*	1.82
$B \times e_{Pr} \times s$	4	1.26	0.62	1.99	0.25	0.04	1.72	0.38
$e_w \times e_{Pr} \times s$	8	1.62	1.44	2.11*	0.65	1.04	1.58	0.48
$B \times e_w \times e_{Pr} \times s$	8	0.97	1.13	0.83	1.42	0.83	0.21	2.99**

Table 5.23: ANOVA results for RANK scores in experiment 2

5.3 Results of Simulation Experiment 3

The third and final simulation experiment investigates the influence of four well-known non-idealities that may surface during practical applications: omission of criteria, violations of independence, shifts in the reference levels, and insufficiently detailed modelling of the partial utility functions. These non-idealities were previously analysed by Stewart [86], although not in a scenario-based environment. The aim here is therefore to contrast those results with results obtained from a scenario-based approach, and to provide a benchmark against which to judge the other non-idealities specific to this study. There are five main effects, four of which correspond to the non-idealities: the number of omitted criteria (o), the mixture of criteria used to induce independence violations (γ), the shift in the reference level (e_r), and the number of segments used to model the partial utility functions (v). The fifth and final main effect relates to the number of scenarios used in the analysis (s). The values assumed by the simulation parameters are indicated in the following table.

Parameter	Values
v	1, 2, 4
e_r	0, 0.1
γ	0, 0.1, 0.3
o	0, 1, 2
s	3, 5, 10

Table 5.24: Parameters for experiment 3

5.3.1 ANOVA Results

Full ANOVA tables for experiment 3 can be found at the end of section 5.3 in tables 5.34 and 5.35. The ANOVA results show a fairly high degree of consistency between the significant factors affecting the POSN and RANK scores, albeit with some exceptions. The RANK scores are generally associated with larger F-statistics than the POSN scores, so that the non-idealities have a more significant effect on the relative performance of the alternative selected by the models. The omission of criteria and violations of preferential independence are the two most significant effects. The F-statistics for the effect of independence violations are more consistent over the models used, generally being around 200 for the POSN scores and 300 for the RANK scores. Although the F-statistics corresponding to the effect of omitting criteria are generally slightly lower, they are also more variable, and are particularly large for the additive and expected Z models. The effect of the number of segments used to model the utility functions is also significant at the 0.5% level in all the models, with F-statistics generally of the same order as the omission of criteria. There is less agreement regarding the significance of the effect of shifts in the reference levels. The shift effect is significant at the 0.5% level in both POSN and RANK scores in only three of the models, while the other four models have conflicting POSN and RANK results. The

effect is also on occasion identified as non-significant.

There is generally only little consistency for the second-order effects. All the second-order effects are identified as significant by at least one model. However, there are only three effects that are *consistently* significant: between the number of segments and reference shifts, between the number of segments and independence violations, and between the number of scenarios and criteria omissions. Although again there are several higher-order interactions of varying significance levels, they are generally inconsistent over model types and the type of score used, and are therefore not of broader interest.

5.3.2 Characterisation of Effects

The four non-idealities considered in this simulation must in some sense cause a deterioration in the quality of the rank approximations. The comparison of the aggregate results obtained in this simulation with the appropriate ‘ideal’ results obtained from simulation 1 is an informative introduction to the analysis, answering the question as to what extent these non-idealities affect the results in the broadest manner possible. As for the previous simulation experiment, we must consider only certain cases from the first simulation in order to ensure that the results differ only with respect to the presence or absence of the appropriate non-idealities. Table 5.25 shows the grand average POSN and RANK scores for each model under simulation 3 and the part of simulation 1 using low τ , low λ , high B and $\kappa = 3$. Table 5.26 shows similar results for the rank correlations.

Model	POSN _{ideal}	POSN _{non-ideal}	RANK _{ideal}	RANK _{non-ideal}
Additive	1.44	3.56	1.47	3.25
Ignore	5.95	6.25	5.25	5.67
Expected Z	3.30	4.56	2.74	4.00
Scenario R	5.67	6.11	5.37	5.80
Scenario P	5.21	5.81	4.99	5.57
Scenario ML	4.28	4.92	4.25	4.79
Scenario F	5.92	6.32	5.32	5.94

Table 5.25: Average POSN and RANK scores for experiment 3 (non-ideal) and corresponding experiment 1 (ideal) problem contexts

Model	SRC_{ideal}	$SRC_{non-ideal}$
Additive	0.94	0.62
Ignore	0.27	0.22
Expected Z	0.65	0.46
Scenario R	0.34	0.26
Scenario P	0.37	0.29
Scenario ML	0.48	0.38
Scenario F	0.30	0.23

Table 5.26: Average rank correlations for experiment 3 (non-ideal) and corresponding experiment 1 (ideal) problem contexts

The most striking feature of both the scores and correlations is the different magnitudes of impact that the non-idealities have on the different models. The additive model is particularly hard-hit, with deteriorations of approximately 2 rank positions in both the POSN and RANK rank orders. Large deteriorations of around 1.3 positions are also observed in the expected Z model. The rank correlation of each model deteriorates by around 30%. The scenario-based models are subject to far smaller deteriorations, generally between 0.5 and 0.6 positions, as a result of the non-idealities. The model ignoring uncertainty is the least influenced by non-idealities, with deteriorations of around 0.3 positions. There are two possible explanations for the different degrees of deterioration. Firstly, the scenario-based models and the model ignoring uncertainty may genuinely be more robust to the non-idealities than the additive and expected Z models. Otherwise, the poor general performance of the former models may imply a diminished sensitivity to further non-idealities. We return to this question at a later stage; at this point it is sufficient to mention that the robustness of a model seems rather a moot point when the average performance is only marginally better than a random guess.

Number of Segments Modelling the Partial Utility Functions

The importance of sufficiently detailed modelling of the partial utility functions is not immediately clear from the simulation results. Table 5.27 shows the POSN and RANK scores for different values of v , the number of piecewise linear segments used. Fairly significant deteriorations of between 0.4 and 0.5 are experienced in the additive and expected Z models when using a single linear segment relative to using four segments. The rank probabilities show less substantial changes, with the increasing of segments from one to four resulting in only a 10% improvement in the frequency of the top ranks coinciding. The deteriorations when reducing the number of segments from four to two are generally similar to those obtained by reducing the number of segments from two to one. However, deteriorations of a similar nature are extremely limited in the other models, generally around 0.15 for the POSN scores and 0.3 for the RANK scores. Furthermore, the deteriorations only occur when moving from four to two segments. No further deteriorations are exhibited when further reducing the

number of segments to one.

These results conflict quite markedly with those obtained by Stewart in [86], where in particular it was observed that using only one segment led to extremely poor results. The most likely reason for these discrepancies lies in the fixing of the shape of the partial utility functions in this simulation study. All four forms of utility functions were simulated by Stewart, while only the first form is used here. Furthermore, that shape turns out to be conducive to linear interpolation between worst and best. The functional form generates τ_j and λ_j from the same interval $[0.15, 0.4]$, so that the utility function lies below the linear segment below the reference point and above the linear segment above the reference point. This is in contrast to cases 2 and 3, where the different intervals for generating the τ_j and λ_j imply a utility function wholly below or above the linear segment respectively. Also, the utility functions have been simulated with the assumption that the slope of the utility function is greater in the domain of losses i.e. below the reference point, implying that the utility function is relatively more non-linear below the reference point. For the utility functions belonging to case 1, the low reference level means that this highly non-linear portion of the utility function is small relative to the utility functions generated according to case 4, for which the majority of the domain is represented by the more strongly non-linear convex utility function.

As a result of these functional considerations, the simulation results obtained here should not be interpreted as contradicting those obtained by Stewart. More specifically, they contribute by providing supplementary evidence that results are *not* severely downgraded by the use of a single segment, providing that the underlying preferences are represented by the utility functions in case 1. If underlying preferences are represented by any of the other cases, then, following the results of Stewart, results are likely to be considerably degraded. This appears particularly true in the case of convex or concave utility functions. The practical implications of these findings are hampered somewhat by the fact that we need to know the form of the underlying utility functions before we can make a judgement about the necessary level of approximation required for adequate representation. This would defeat the purpose of the approximation. We may, however, elicit the reference level and the preference for the reference level relative to the worst and best outcomes using questions inspired by some sort of thermometer-scale. In doing so it may be possible to classify the utility function to one of the four functional forms. On the basis of this classification, it might then become more apparent to what extent a multi-segment approximation is required.

v	POSN			RANK		
	1	2	4	1	2	4
Additive	3.74	3.56	3.38	3.55	3.22	2.99
Ignore	6.30	6.30	6.15	5.80	5.65	5.55
Expected Z	4.68	4.63	4.36	4.28	3.99	3.75
Scenario R	6.15	6.17	6.02	5.90	5.83	5.66
Scenario P	5.85	5.85	5.72	5.66	5.59	5.44
Scenario ML	4.96	4.99	4.80	4.89	4.83	4.66
Scenario F	6.33	6.38	6.25	6.05	5.98	5.78

Table 5.27: Effect of number of segments modelling the partial utility functions (v)

Shifts in Reference Level

The deteriorations due to shifts in the reference levels are comfortably the smallest of the non-idealities, and in the scenario models and the model ignoring risk exert a negligible influence on the results. Although larger deteriorations are experienced in the additive and expected Z models, the magnitude of these effects remains at most moderate and is unlikely to be of any kind of critical importance. Table 5.28 gives the POSN and RANK scores for different shifts in the reference levels, e_r .

There is one two-way interaction between reference shifts and the number of segments that, although of only moderate magnitude, is systematic enough to motivate further investigation. Table 5.29 displays the deteriorations in scores due to shifts in the reference level categorised by the number of segments used. The deteriorations due to shifts in the reference levels are more prominent when more linear segments are used to approximate the partial utility functions. This is observed in all the models, but due to the generally negligible reference shift effect, it is only in the additive and expected Z models that the interaction is of any importance. There is in fact a simple and uninteresting reason for the interaction: the closer the approximations to the utility functions are, the more dependent they become on the precise form of the utility functions, therefore becoming more sensitive to changes in the utility functions. As a result the four-segment approximation picks up the reference shift more accurately than the two-segment or one-segment approximations. In fact the interpolation between worst and best attribute values using a single linear segment makes no use of the underlying form of the utility function, and hence is not influenced by reference levels or shifts thereof. Table 5.29 therefore excludes the case of a single linear segment.

	POSN		RANK		
	e_r	0	0.1	0	0.1
Additive		3.43	3.69	3.08	3.42
Ignore		6.23	6.27	5.71	5.62
Expected Z		4.45	4.66	3.93	4.08
Scenario R		6.10	6.12	5.84	5.75
Scenario P		5.78	5.84	5.60	5.53
Scenario ML		4.89	4.94	4.80	4.78
Scenario F		6.32	6.32	5.97	5.91

Table 5.28: Effect of reference level shifts (e_r)

	POSN		RANK		
	v	2	4	2	4
Additive		+0.26	+0.48	+0.32	+0.68
Ignore		+0.00	+0.14	-0.15	-0.09
Expected Z		+0.19	+0.45	+0.11	+0.36
Scenario R		+0.00	+0.10	-0.13	-0.13
Scenario P		+0.05	+0.14	-0.18	-0.03
Scenario ML		+0.00	+0.16	-0.10	+0.04
Scenario F		+0.00	+0.08	-0.17	+0.00

Table 5.29: Effect of interaction between the number of segments (v) and reference level shifts (e_r)

Independence Violations

Some fairly substantial deterioration of results is possible as a result of independence violations, particularly when these violations become large. A 10% mixing of criteria results in increases in the average scores of between 0.1 and 0.2, and so can be considered marginal and of limited practical interest. However, increasing the mixing of criteria to 30% results in further deteriorations of between 0.35 and 0.45 positions. Large violations of independence may therefore have quite significant impacts, up to 0.7 positions in this case. The magnitude of the effects is remarkably consistent over models in both the POSN and RANK scores. Furthermore the deteriorations in POSN and RANK scores are also very similar. In contrast, violations of independence have only modest effects on the rank probability matrix. Severe violations lead to decreases of up to 10% in the frequency that the true best alternative is the same as the alternative proposed by the model.

	POSN			RANK			
	γ	0	0.1	0.3	0	0.1	0.3
Additive		3.33	3.46	3.89	3.07	3.14	3.55
Ignore		6.11	6.23	6.42	5.47	5.61	5.92
Expected Z		4.26	4.46	4.94	3.73	3.87	4.41
Scenario R		5.88	6.04	6.41	5.55	5.71	6.12
Scenario P		5.60	5.74	6.09	5.33	5.46	5.90
Scenario ML		4.70	4.85	5.20	4.57	4.70	5.12
Scenario F		6.11	6.25	6.60	5.70	5.86	6.25

Table 5.30: Effect of violating independence (γ)

Omission of Criteria

The number of criteria omitted from the analysis is modelled by the parameter o . Table 5.31 shows the POSN and RANK scores when different numbers of criteria are omitted. An interaction between the number of omitted criteria and the number of scenarios selected is shown in table 5.32, which displays the increases in POSN and RANK scores relative to the case when no criteria are omitted, when different numbers of scenarios are selected.

The omission of even one criterion leads understandably to a fairly large deterioration in the quality of the results, for both the additive and the expected Z model. The additive model is particularly badly affected, with a deterioration of 1.6 in the POSN ranking and 0.8 in the RANK ranking. In such circumstances the probability that the true best alternative is among the first three model ranks drops from an acceptable 85% to 65%, and the best alternatives coincide in less than 30% of simulations. The deteriorations are much less pronounced for the other models, and in fact for the POSN and RANK scores are generally quite marginal at around 0.15. Any changes in the rank probability matrix are negligible. Again, this may be interpreted as an indication of the relative robustness of the scenario-based models towards criteria omissions, but is also explained as the effect of decreasing marginal sensitivity towards non-idealities when model performance deteriorates, and therefore as a reflection of the relatively poor performance of the scenario models.

The progressive deterioration is fairly linear in the number of criteria omitted, so that results are proportionately worse when two criteria are omitted. However, the relatively greater degeneration of the additive model means that, when two criteria are omitted, the performance of this model is closer to the other models than before. In particular the additive model outperforms the expected Z model by only 0.3 positions in the POSN and RANK scores when 2 criteria are omitted. Under these conditions, the probability that the top three model ranks contain the true best alternative is generally poor: 52% in the additive model, 47% in the expected Z model, and 51%

in the scenario model selecting the ten most likely scenarios. The relatively better performance of the scenario model provides some evidence of genuine robustness to criteria omissions rather than any diminished sensitivity. The practical implications are obvious: the additive model is not particularly robust to omissions of criteria, and due care needs to be taken to ensure that such omissions do not occur.

A clear interaction exists between the effects of the number of scenarios selected and the number of criteria omitted, although the magnitude of the effect is not particularly large due to the limited effect of criteria omissions in the scenario models. Nevertheless, the interaction is of interest for its systematic behaviour rather than any crucial importance to the integrity of results. Essentially, the deteriorations due to criteria omissions are greater when more scenarios are included in the analysis. In fact, if only 3 out of the 50 scenarios are considered, omitting 2 criteria causes only a 0.1 position deterioration in results. This is taken as further evidence of the diminishing sensitivity to non-idealities as performance worsens. If 10 scenarios are considered, the magnitude of the deteriorations is approximately equal to that of the expected Z model.

	POSN			RANK			
	<i>o</i>	0	1	2	0	1	2
Additive		2.18	3.74	4.75	2.41	3.22	4.12
Ignore		6.19	6.26	6.31	5.54	5.66	5.79
Expected Z		4.06	4.58	5.02	3.52	4.01	4.49
Scenario R		5.98	6.12	6.23	5.73	5.79	5.87
Scenario P		5.67	5.82	5.94	5.48	5.56	5.65
Scenario ML		4.70	4.91	5.13	4.66	4.78	4.94
Scenario F		6.21	6.31	6.44	5.85	5.93	6.03

Table 5.31: Effect of omitting criteria (*o*)

	<i>s</i>	POSN		RANK	
		omit 1	omit 2	omit 1	omit 2
Scenario R	3	+0.10	+0.11	+0.05	+0.06
	5	+0.11	+0.15	-0.04	+0.01
	10	+0.22	+0.50	+0.14	+0.34
Scenario P	3	+0.10	+0.11	+0.00	+0.06
	5	+0.13	+0.22	+0.04	+0.08
	10	+0.19	+0.47	+0.19	+0.37
Scenario ML	3	+0.09	+0.17	+0.06	+0.10
	5	+0.21	+0.35	+0.07	+0.20
	10	+0.32	+0.77	+0.22	+0.52
Scenario F	3	+0.05	+0.07	+0.11	+0.17
	5	+0.09	+0.19	+0.06	+0.14
	10	+0.17	+0.41	+0.05	+0.24

Table 5.32: Effect of interaction between the number of selected scenarios (*s*) and omission of criteria (*o*)

General Robustness of Results

The discussion of relative robustness concludes the presentation of the results of simulation experiment 3. From the previous sections we may infer that the accuracy of results will deteriorate as the number of piecewise segments used decreases, reference shifts increase, violations of independence increase, the number of omitted criteria increases and the number of scenarios considered decreases. Table 5.33 shows the probabilities of locating the true best alternative in the top two model ranks under conditions of varying difficulty.

	Worst	Average	Best
Additive	0.34	0.56	0.92
Ignore	0.21	0.26	0.32
Expected Z	0.30	0.41	0.55
Scenario R	0.23	0.30	0.36
Scenario P	0.27	0.34	0.40
Scenario ML	0.33	0.44	0.53
Scenario F	0.22	0.29	0.35

Table 5.33: Probabilities that the true best alternative is placed first or second in the model rank order

The movement away from the superior robustness of the additive model that was first observed in experiment 2 is continued here. In fact although in experiment 2 the relative deteriorations of some of the scenario-based models were still greater than the additive model, this is no longer the case. The additive model experiences

the most severe absolute and relative deteriorations, and under extremely non-ideal conditions the relative superiority of both the additive and the expected Z models over the scenario ML model is extinguished. As in both previous experiments, the superiority of the additive model in ideal or close-to-ideal conditions is evident, while the performance of the expected Z models is no better than the best scenario model. However, the results of both models deteriorate relatively quickly as conditions move away from ideal. The average probability values lie closer to the worse-case value than the best-case value, in contrast to the scenario models which are generally symmetrical. In other words, the frequency distribution of this particular success measure is symmetrical or skewed very slightly to the right in the scenario models and skewed to the right for the additive and expected Z model.

The reasons for the shifting robustness behaviour can again be explained by considering two types of robustness relationships. Experiment 1 supported the sensitivity of the scenario models to the effect of attribute variability, which is excluded from this experiment. We could therefore expect the scenario models to *ceteris paribus* be relatively more robust in experiment 3. Then, the discussions in the previous section have shown that the additive model, and to a lesser extent the expected Z model, are more sensitive to the new non-idealities introduced into experiment 3, with the exception of violations of independence. These two relationships conspire to cause the described changes in robustness. In fact since the scenario ML model was shown in experiment 1 to be relatively more robust to attribute variability than the other scenario models, the exclusion of that effect leaves the relative robustness constant across the scenario models. A secondary robustness relationship is thus that all scenario models are similarly robust to the non-idealities specific to experiment 3.

5.3.3 Summarised Inferences

The following set of summarised inferences are drawn based on the results of simulation experiment 3:

- The use of one linear segment is not particularly harmful provided the underlying partial utility functions have the functional form characterised by case 1 i.e. moderately but not extremely low τ_j and λ_j , and gradients steeper below the reference level than above it.
- Deteriorations due to shifts in the reference levels are no more than moderate for even fairly large shifts.
- Violations of preferential independence may cause fairly substantial deteriorations if the violations are of sufficient magnitude. Moderate violations of independence are unlikely to cause practically significant deteriorations.
- The omission of criteria results in very substantial deteriorations in the additive and expected Z models, but only marginal deteriorations in the other models.

The deterioration is linear in the number of criteria omitted for the simulated range of two criteria.

- Although the omission of criteria does not result in large deteriorations, those deteriorations do become more pronounced as the number of scenarios selected increases.
- There is further evidence of the robustness of the scenario models and the model ignoring risk to the simulated non-idealities. These models experience relatively less deterioration than the additive or expected Z models.
- The robustness of the scenario models is independent of selection strategy.

University of Cape Town

Effects	DoF	Additive	Ignore	Expected Z	Scenario R	Scenario P	Scenario ML	Scenario F
v	2	176.93**	22.90**	126.26**	20.44**	17.63**	40.81**	15.16**
e_r	1	261.63**	4.09*	138.84**	0.89	11.00**	9.23**	0.15
γ	2	452.67**	79.12**	506.36**	228.68**	197.07**	257.76**	189.44**
o	2	8820.25**	12.11**	947.18**	49.88**	55.60**	181.80**	39.75**
s	2	0.21	1.21	1.14	813.38**	1061.33**	1780.42**	773.06**
$v \times e_r$	2	69.46**	5.63**	51.34**	3.63*	3.11*	7.96**	3.64*
$v \times \gamma$	4	6.33**	0.69	4.73**	0.91	1.63	1.93	3.14*
$e_r \times \gamma$	2	1.36	4.21*	5.36**	0.25	0.40	0.21	0.01
$v \times o$	4	1.65	0.80	4.33**	1.18	1.54	0.98	3.65*
$e_r \times o$	2	15.30**	0.19	0.65	2.37	0.74	0.11	2.68
$\gamma \times o$	4	10.59**	1.42	2.49*	0.66	0.57	0.50	3.05*
$v \times s$	4	0.50	1.16	1.09	1.27	1.83	2.56*	2.45*
$e_r \times s$	2	1.17	0.74	0.34	3.30*	1.10	0.98	3.61*
$\gamma \times s$	4	0.88	2.02	1.66	3.31*	2.40*	1.08	3.32*
$o \times s$	4	1.79	0.27	0.81	12.64**	9.95**	32.73**	7.80**
$v \times e_r \times \gamma$	4	0.84	1.27	1.05	2.31	0.46	1.29	0.53
$v \times e_r \times o$	4	4.09**	2.46*	1.42	1.23	1.05	1.15	0.37
$v \times \gamma \times o$	8	1.78	0.81	1.04	0.74	0.68	0.71	1.03
$e_r \times \gamma \times o$	4	1.06	0.82	1.25	0.50	1.01	0.88	1.65
$v \times e_r \times s$	4	1.95	0.27	0.52	1.77	0.90	0.89	1.83
$v \times \gamma \times s$	8	0.75	0.42	0.96	1.42	0.26	1.17	0.98
$e_r \times \gamma \times s$	4	0.55	0.50	0.26	0.39	0.95	0.97	0.71
$v \times o \times s$	8	0.33	0.17	0.37	0.32	1.45	1.19	0.82
$e_r \times o \times s$	4	2.10	2.70*	2.35	1.08	0.95	0.57	0.79
$\gamma \times o \times s$	8	1.46	0.37	0.46	0.27	0.29	0.69	2.13*
$v \times e_r \times \gamma \times o$	8	0.35	2.17*	0.99	0.63	0.40	1.40	1.26
$v \times e_r \times \gamma \times s$	8	0.97	1.43	1.35	1.08	1.46	1.40	1.19
$v \times e_r \times o \times s$	8	0.72	0.66	0.81	1.81	1.10	1.38	1.23
$v \times \gamma \times o \times s$	16	0.48	1.38	0.88	0.88	1.29	0.78	0.84
$e_r \times \gamma \times o \times s$	8	0.97	1.32	0.64	0.47	2.03*	0.58	0.82
$v \times e_r \times \gamma \times o \times s$	16	0.60	0.72	0.58	0.77	0.98	1.00	0.94

Table 5.34: ANOVA results for POSN scores in experiment 3

Effect	DoF	Additive	Ignore	Expected Z	Scenario R	Scenario P	Scenario ML	Scenario F
v	2	608.53**	56.22**	400.26**	56.25**	52.59**	61.81**	67.90**
e_r	1	642.10**	25.14**	106.16**	20.20**	15.79**	1.14	8.37**
γ	2	497.78**	196.67**	738.46**	317.59**	370.20**	345.03**	278.95**
o	2	5515.10**	54.34**	1355.78**	17.18**	29.54**	78.39**	28.23**
s	2	2.27	0.29	0.16	991.76**	1463.16**	1914.61**	544.11**
$v \times e_r$	2	207.86**	2.78	50.03**	5.80**	8.95**	4.86*	7.53**
$v \times \gamma$	4	13.76**	0.08	12.72**	1.00	2.80*	2.47*	3.46*
$e_r \times \gamma$	2	10.21**	0.31	3.26*	1.54	0.09	1.01	0.50
$v \times o$	4	2.22	1.40	2.77*	1.88	1.92	4.30**	4.22**
$e_r \times o$	2	13.44**	1.54	1.26	2.29	7.48**	1.23	3.09*
$\gamma \times o$	4	2.64*	1.06	1.18	1.98	0.16	0.52	2.28
$v \times s$	4	0.54	0.88	0.59	1.53	2.06	3.51*	4.05**
$e_r \times s$	2	0.52	0.70	0.63	5.89**	0.32	2.07	3.01*
$\gamma \times s$	4	1.12	2.77*	1.06	1.87	10.62**	4.38**	0.75
$o \times s$	4	1.01	1.51	1.38	9.97**	10.61**	16.72**	1.44
$v \times e_r \times \gamma$	4	3.31*	1.44	2.33	0.98	1.65	0.50	0.80
$v \times e_r \times o$	4	3.34*	0.94	1.09	0.71	0.49	0.43	0.31
$v \times \gamma \times o$	8	1.65	1.18	1.62	1.66	0.73	0.61	1.04
$e_r \times \gamma \times o$	4	2.35	0.91	0.53	1.46	2.17	0.42	1.13
$v \times e_r \times s$	4	1.05	0.67	0.95	2.02	1.41	1.01	1.08
$v \times \gamma \times s$	8	0.27	0.95	0.72	0.97	1.09	0.36	3.05**
$e_r \times \gamma \times s$	4	0.41	1.62	0.92	1.15	0.77	2.98*	0.06
$v \times o \times s$	8	0.96	1.06	0.97	1.15	1.36	0.92	1.50
$e_r \times o \times s$	4	1.01	2.72*	1.43	3.12*	0.93	0.98	0.78
$\gamma \times o \times s$	8	0.71	0.66	1.07	0.95	0.91	1.09	0.98
$v \times e_r \times \gamma \times o$	8	1.16	1.49	0.37	1.57	0.69	1.52	0.53
$v \times e_r \times \gamma \times s$	8	1.85	0.78	0.65	0.52	1.10	0.94	2.16*
$v \times e_r \times o \times s$	8	0.43	0.79	0.61	1.43	1.16	1.08	0.54
$v \times \gamma \times o \times s$	16	0.54	1.35	1.76*	0.82	1.18	0.96	1.47
$e_r \times \gamma \times o \times s$	8	0.30	1.08	0.30	1.17	1.13	1.14	0.52
$v \times e_r \times \gamma \times o \times s$	16	0.47	1.20	0.67	0.93	1.08	0.97	0.84

Table 5.35: ANOVA results for RANK scores in experiment 3

5.4 Post-hoc Reflections on the Simulations

It has been stressed previously that each simulation is an exercise in abstracting reality to a point where it retains its fundamental features while allowing for a meaningful interpretation of its simplified form. This process involves a series of choices that, while furthering the aims of a particular experiment, may be less suitable for others. This incongruity may become especially clear, as was the case here, where post-hoc modifications are attempted in response to certain aspects of the simulation results. The results of the simulation experiments presented in this chapter suggested a course of action that was simply not supported by the constructed simulation structure. In the interests of rigour and as an aid to future research we present a brief discussion of these problems.

In light of the poor performance of the scenario-based models we were led to consider whether this was a general result, or whether there was *any* way of selecting scenarios that would result in a better approximation to the true rank order. In some independent work, Stewart (personal communication, paper in preparation) considered a selection strategy making use of the 5th and 95th percentiles of each alternative's performance on each criterion. The 5th percentiles are grouped into a single scenario, representing in some sense a 'bad' scenario, while the 95th percentiles are grouped into a corresponding 'good' scenario. These scenarios are considered in conjunction with what we have termed the status quo scenario; that is the 'average' set of evaluations to which are added the scenario-based perturbations, giving a set of three scenarios. The 'good' and 'bad' scenarios are artificial in the sense that they are not plausible future states of the world, as the scenarios defined for the purposes of our simulation were. In fact, as they are constructed from several distinct states of the world they ultimately embody different and necessarily conflicting chains of causal reasoning, which does present some qualitative challenges. Nevertheless, a similar approach is often employed in scenario planning practice, and the results of some preliminary experiments conducted with this different simulation structure showed considerable promise. Using slightly different percentiles, as well as including further scenarios, gave similarly impressive results. On the basis of these preliminary results an attempt was made to replicate them using the simulation structure discussed in chapter 4.

However, in the original simulation structure that was used the alternatives were constructed to be of very similar robustness. That is, the additive perturbations used to model the modifying effect of different scenarios on the 'status quo' evaluations Z_{ij1} are all generated uniformly between $[-b, +b]$, where b is the parameter of the simulation modelling attribute variability. Nevertheless, the value of the parameter is constant over alternatives, with the result that the range between the best and worst possible performance in each alternative is very similar for a particular criterion. This

similarity implies that the preference orders derived from each of the ‘good’, ‘average’ and ‘bad’ scenarios are also very similar in nature: little new information is captured by considering multiple scenarios. While the multiplicative scenario effect C_{jk} does cause *some* differences in robustness to appear, it is not designed to model robustness among alternatives; the changes in relative ranges of performance are difficult to determine and control. In contrast, the simulation structure used by Stewart explicitly assigns different levels of robustness to each alternative within each criterion.

Using the original simulation structure, the post-hoc simulation of the ‘quartile’ scenario selection method did in fact produce excellent results – on average 1 to 1.5 rank orders better than the previous best scenario-based model. However, the success appears to be due to the attribute values in the status quo scenario closely resembling the true expected attribute value, which has been shown to be a profitable approach. Inasmuch as it is not possible to trace the definite effects of this apparently promising selection strategy, the manner in which the attribute evaluations are constructed must be considered a weakness of the simulation. Strictly speaking the results of the three simulation experiments should be interpreted within the context of alternatives exhibiting similar levels of robustness. It does seem important, however, that some research be directed towards the evaluation of this artificial creation of ‘quartile’ scenarios. Future research could also examine to what extent allowing for different levels of robustness among the set of alternatives influences the results obtained here. Both these research objectives could be achieved, for example, by simply modifying the existing simulation structure by requiring that the additive perturbations B_{ik} are generated uniformly between $[-b, +b]$ where b is no longer constant over alternatives but is a function of at least i . The precise functional form would dictate the extent to which the alternatives differ in terms of robustness.

Chapter 6

Conclusions and Future Research

6.1 Insights from the Simulation Study

The previous chapter has identified the more specific results relating to the simulation experiments and summarised them as stylised inferences. In this section, we attempt to provide a more general set of conclusions by synthesising the information contained in the three simulations with reference to the three main research questions identified in chapter 4. It becomes meaningful to separate the parameters of the simulations into contextual parameters, that is those effects that essentially exist outside of the *practical* modelling of the decision problem, and non-idealities, those effects which relate to the physical building of the decision model. Under the heading of contextual effects we can classify all the parameters of experiment 1: reference level, utility at the reference level, attribute variability and multivariate risk proneness. We can label as non-idealities all those parameters specific to experiments 2 or 3: assessment errors of criterion weights and scenario probabilities, number of segments modelling the partial utility functions, independence violations, reference shifts and criterion omissions.

Dominance of the Additive Model

1. One of the strongest messages emerging from the simulation results is the generally excellent performance of the additive model: that is, the model considering the full set of 50 scenarios in an additive approximation to the global utility function. It appears that very little is lost by considering an additive functional form relative to the more complicated multiplicative form. In particular, the additive model may be used with confidence if the desired output is a shortlist of 2 or 3 alternatives. These results are not new; similar ones were reported by Stewart [85], although in the simulations performed here it is possible to obtain an idea of the magnitude of the superiority of the additive model. A key point emerging from our simulations, however, is that the simplification of the *preference* model appears to cause far less of a deterioration in results than does the simplification of the *stochastic* structure of the model. In most cases the additive model very substantially outperforms even its closest rival,

the expected Z model. The analyst should ensure that the focus of the decision process on the multicriteria aspects of the problem is not to the detriment of the modelling of risk.

2. It is also apparent that the additive model is substantially more robust to changes in the contextual parameters. Specifically, it is unaffected by attribute variability and affected far less than other models by changes to the functional form of the partial utility functions. The additive model therefore performs not only far better, but more consistently than the other models when no non-idealities are present.
3. Although the additive model is considerably more sensitive to the occurrence of non-idealities than the other models, it remains the best-performing model under all circumstances, and the resulting deteriorations are not of serious concern provided that the non-idealities do not become extreme. An important exception is the omission of criteria: omitting even one out of seven criteria causes very substantial deteriorations in the quality of the results and should be closely guarded against. The analyst should ensure that sufficient time is spent focusing the attention of the decision maker on the definition of a complete set of criteria during the problem structuring phases of the decision process. Probing whether all dimensions of the problem have been captured as criteria requires some creative thinking from both the decision maker and the analyst, but might take the form of discussing firstly the ways in which alternatives equally rated on all captured criteria may differ, and then further discussing whether these aspects are important enough to include in the analysis as model criteria.

Plausibility of Simplification Strategies

1. One of the most interesting aspects of the results is the evidence suggesting that estimating the expected attribute values with a moderately high degree of approximation may be a useful simplification approach. In nearly all circumstances the performance of the expected Z model is second only to the additive model in terms of accuracy. In fact the expected Z model performs marginally but consistently better than the best scenario results i.e. when $s = 10$. Given the success of this model, future research might be directed into the development of problem structuring tools capable of closely approximating the true expected attribute values. Such tools would *inter alia* need to take into account well-known heuristics and biases used in the assessment of probabilities.
2. The results of the simplification models are less sensitive to changes in the problem context than the scenario models, so that performance is generally good under a wider range of problem contexts when no non-idealities are present. However, the model is far more sensitive to changes in the contextual parameters than the additive model, with the exception of multivariate risk proneness.

3. Where non-idealities are present, the expected Z model experiences relatively greater deteriorations than the scenario models, although far less than the additive model. The superiority of the expected Z model over the scenario models is therefore less certain when even moderate levels of non-idealities are present, and can become an inferiority as the non-idealities become extreme. Nevertheless, the generally good performance of even this simple model provides some justification and motivation to consider the use of expected values as a valid simplification technique.
4. An equally important message from the simulation results is that if the estimation is poor, then results are severely degraded. This second simplification model, the so-called model ignoring risk, performed extremely poorly under all conditions. Even under optimal conditions with no non-idealities, the accuracy of the model is probably not sufficient for practical purposes. The results of the two models can therefore be taken together as a warning that any estimation of expected values must be done with due care. If such care is not taken, the quality of the results will reflect that.

Danger of Scenario-based Approaches

1. The simple core of the message provided by all three simulations is that the injudicious use of scenarios may lead to results that are unacceptably poor for broader MCDM application. In particular, the general performance of the models using either 3 or 5 actual scenarios is poor regardless of selection policy. Although relatively efficient use is made of the information used, the fact remains that it is only 5-10% of the total attribute information. Some loss of quality therefore seems inevitable, but it is of paramount importance that the loss does not become so large as to jeopardise the feasibility of the model. Yet this is precisely what has happened with the simulated scenario models. It is a fundamental conclusion of this thesis that the danger of applying existing scenario models has been understated in the literature. Nevertheless, it is important to note that we have only simulated some plausible selection policies. Future research should be directed towards the evaluation of other selection policies and other problem contexts. In particular, early investigations into the use of quartiles of performance evaluations to construct three artificial 'good', 'medium', and 'bad' scenarios indicate that this approach might prove fruitful.
2. The poor performance of the scenario-based models is indicative of the problems involved both in the estimation of expected values and the identification of diversity in a multivariate scenario environment. In particular, the simulation results suggest that certain interpretations of a 'diverse' set of scenarios can lead to unacceptably poor results, which leads to an important message of this thesis: for MCDM and scenario planning to be applied in an integrated fashion, an operationally useful basis for selecting diverse scenarios (beyond that provided

by the scenario planning literature) needs to be found that avoids the type of deteriorations experienced here.

3. Although the results of the scenario models are generally poor, when 10 scenarios out of 50 are used the results may be quite good under optimal conditions. The performance of the model selecting the ten most likely scenarios is then very similar in quality to the expected Z model. For practical purposes it is unlikely that as many as ten scenarios are thoroughly considered, but there is at least some evidence that partitioning the criterion space well may bring positive results. Although the performance under five scenarios remains poor, there is evidence that the marginal improvements from the three-scenario model more than justify the effort involved in creating the additional scenarios.
4. The scenario models are highly sensitive to changes in the problem context, and are particularly severely affected by changes in the degree of compensation. Greater levels of attribute variability also understandably result in fairly large deteriorations. However if both the reference levels and attribute variability are favourable, the scenario models perform well even if only five scenarios are considered. The general use of scenario models in the sense considered here can only be justified if reference levels are high and attribute variability is low.
5. Although the scenario models are sensitive to the problem context, they are relatively more robust to the non-idealities than either the additive or expected Z models. Although it is difficult to isolate the extent to which the apparent insensitivity is due to general poor performance, there does appear to be at least some genuine robustness. This provides some indication that if the general performance of the scenario approaches could be improved, they could be more suitable to problems which are poorly structured. Such problems might include hastily constructed analyses, unimportant decisions or unfacilitated analyses. At the present junction, though, the robustness of the scenario models is very much secondary to their poor performance, and should not be overemphasised.
6. In general, the quality of approximation increases with an increasing dependence on probability of occurrence as a method of scenario selection. In fact, the consideration of a diverse (in the context defined here) set of scenarios offers no better approximation to the true rank ordering than a random selection. However, some mitigating factors need to be considered. The use of diverse scenarios offers benefits relating to indirect aspects of the decision that were not simulated. Future simulations might consider that the use of scenario planning illuminates a set of possible scenarios that are not visible to the other models.

Robustness to Non-Idealities

1. The MCDM models are generally fairly robust to the occurrence of moderate levels of non-idealities. The contextual effects that will be present in any de-

cision problem are of a similar magnitude to the effects of non-idealities. The scenario models are most robust to non-idealities, followed by the simplification models and then the additive model. In contrast the additive model is unaffected by the contextual parameters, while the scenario models are highly sensitive. The simplification models occupy an intermediate position, though closer to the scenario models. The scenario models therefore become relatively more attractive as the framework of the decision becomes more problematic, although all the models may be used with some confidence in their abilities to handle such conditions.

2. Assessment errors are generally not particularly influential unless they are extreme. In particular errors in the assessment of scenario probabilities are negligible for a large range of errors. An implication of this robustness is that it is unlikely to be important whether notions of relative likelihood, which have been rejected by scenario planners, are included in a scenario analysis or not. Weight assessment errors cause substantial deterioration in the additive model when the errors become large, but are only moderate in the other models.
3. The only non-idealities causing consistently substantial deteriorations are criterion omissions and independence violations. The omission of criteria exerts comfortably the most significant influence of the non-idealities, and is extremely detrimental to the quality of results in the additive model. Due care needs to be taken to avoid omitting even one criterion. Deteriorations due to violations of independence are only substantial if severe violations occur, and are unique in that they affect the models equally.
4. The use of a single-segment linear partial utility function *is* justifiable if the functional form of the utility function is such that the reference level and the utility at that reference level are moderately low, with gradients steeper below the reference level. The conclusion does not extend to general functional forms though, and should be understood in the context of other results highlighting the deterioration resulting from the use of a single segment. Shifts in reference levels do not cause substantial deteriorations in any of the models.

6.2 Insights from the Literature Survey

The following conclusions are drawn from the literature survey conducted and described in chapters 2 and 3, with particular reference to the objectives outlined in chapter 1.

Identifying Relevant Uncertainty in MCDM

Three broad, non-exclusive categories of uncertainty are recurrent in the literature: risk, imprecision, and vagueness. Risk relates to the possible dependence of the

performance of an alternative on a random future event. In such a case the physical randomness of the decision problem prevents the performance of the alternative from being known with certainty. Imprecision relates to doubts about the accuracy of physical quantities, beliefs or preferences, and arises from the limited discriminatory power of the human mind. Vagueness relates to doubts arising from the limited amount of knowledge available at any point in time. A fundamental aspect of the decision problem is that we must draw boundaries around the problem to be modelled: vagueness relates to the uncertainties involved in constructing this boundary.

Value Function Methods

Multiattribute utility theory is the most comprehensive treatment of risk available to MCDM practitioners. The axiomatic foundations provide a strong base for both practical and theoretical work. However, strict adherence to the axiomatic requirements of MAUT often necessitates the use of unwieldy models that are arduous to construct. If we wish to incorporate the axiomatic violations that are known to occur in practice i.e. to use a more general set of axioms, then the complexity of the model increases further. Yet the theoretical advancement of more complex methods is often in conflict with the limited consequences of using simpler models that ignore the violations. From a practical perspective, the axioms of MAUT should be interpreted constructively rather than in a strictly normative sense. Although for the integrity of the model it is important that the preferences of the decision maker are in agreement with the axioms, the agreement should not be applied too strictly, but rather used as a goal which is worked towards and as a basis for discussing the tradeoffs at hand.

Outranking Methods

The lack of an integrated treatment of risk in the outranking methods is indicative of the little attention the problem has received. Most of the surveyed approaches hinge on the ability of the DM to make judgements about probability distributions defined over attribute values. Even if the serious question of whether such judgements are practically feasible is ignored, the methods seem to neglect the real aim of the decision process: learning about the nature of preferences and the decision problem at hand. The outranking methodology appears secondary to the treatment of risk, and as such the models appear unbalanced and poorly integrated. The construction of a distributive outranking relation is a far more methodologically attractive option. Although it is considerably more complex, the model possesses the admirable quality of postponing the aggregation over future outcomes, in contrast to the other outranking models and even the MAUT model. The integration of scenario planning techniques with the distributional outranking model should be of considerable interest, based on their common view of outcome aggregation and the need for practical simplification.

Metric Methods

The treatment of risk in a goal programming framework has been largely borrowed from similar models in the older and more well-established linear programming area. Most of the focus has been on chance-constrained models that set goals on the probability that constraints are satisfied. These models are not well suited to MCDM, where there are no strict constraints. In such circumstances, goals must be specified both on the desired level of performance and the desired probability of achieving such a level of performance. The interrelationship between the two aspirations introduces an undesirable complication that will in most cases only serve to confuse the process. Furthermore the consideration of preferences over probability distributions leads to the same practical problems as encountered in the outranking methods. An almost unexplored but seemingly more attractive MCDM model is an extension of the recourse ideology to the GP context, obtained by the minimisation of the expectation of a weighted sum of deviations. The complications stemming from the dual aspirations in the chance-constrained model are thus avoided in this approach. Research should be directed into the methodological and algorithmic feasibility of such a model.

Integrating Scenario Thinking with MCDM

MCDM can appear, both to outsiders and those working in the field, as a house divided against itself; and in many respects this appearance is a fundamental obstacle to the widespread use of MCDM tools. This carries through to the incorporation of risk, where little in the way of a unified approach is available. To this end we have proposed several models making use of scenario planning, a well-established and popular tool for incorporating risk into high-level decision making. We have shown that a scenario-based approach to incorporating risk can be integrated with any of the value function, outranking, or metric MCDM methods. Although no claim is made that a scenario-based approach is universally applicable, it does make some contribution towards an integrated and unified framework for decision making.

A scenario-based approach to MCDM could view the context of each scenario as a deterministic MCDM problem making up a super-MCDM problem, equivalent to the consideration of the scenarios in the second level of the objectives hierarchy. Aggregation over the future outcomes is therefore postponed as far as is possible. The facilitator should use some basic cross-checks to identify whether or not the extra qualitative information illuminated by the construction of scenarios has resulted in scenario-dependent preferences. Some discretion would be required as to whether any differences are sufficient to warrant the increased complexity of a model incorporating changing preferences. Future research would need to consider both the occurrence and effects of changing preferences. The former can only be answered in a practical setting, while the latter might quite easily be tested in a simulation environment.

The nature of the aggregation over scenarios itself is quite controversial. Although scenario planners often advocate 'no aggregation', the implicit message is often that of maximin aggregation, which in many practical applications is overly pessimistic. The simulation results indicate that it is not a good idea to consider the scenario approach as a three- or five-point approximation to expectation. Future research therefore needs to consider softening the notions of robustness to be applicable to the highly conflicting problem contexts found in MCDM. However, for MCDM to be applicable to broader scenario planning, a fundamental adaptation is for the construction and consideration of results within each of the scenarios to assume primary importance. Under such conditions the global aggregation should be considered of secondary importance, and several different forms may be used in a single analysis. The most apparent advantage of using MCDM in a scenario planning context is the formality and structure that is provided to the decision problem in terms of articulating preferences and available tradeoffs.

6.3 Themes for Future Research

In the earlier sections of this chapter and as a footnote to the literature survey in chapter 3 we made mention of several future research questions. In this penultimate section we draw from these questions to select a number of themes for future research.

Development of Problem Structuring Tools

A successful structuring of the problem context to be considered is important in many disciplines; nevertheless those structuring tools currently used in MCDM must adapt to the specific needs of MCDM in order to be relevant. Our simulation results highlighted two areas that might be considered from a problem structuring point of view:

1. The omission of criteria is critical to the integrity of results. In order to avoid omitting any material criteria from the scope of the analysis, the analyst needs a set of structuring approaches at his or her disposal that can be used to probe the preferences of the decision maker at a very early stage of the analysis in order to ensure that the described value tree is in fact complete. Future research could be directed into the qualitative and quantitative structuring tools that might be used.
2. Making use of expected attribute values provides good results in many instances, but it is imperative that the expected values be closely approximated. Future research might consider firstly exactly what qualifies as 'closely approximated' i.e. to extend the results provided in our simulations by considering more than just the two degrees of approximation. Then, the development of problem structuring tools capable of closely approximating the true expected attribute values might be considered.

Unified Risk Treatment in the Outranking and Metric Methods

Unlike the value function methods, where one risk treatment (MAUT) dominates the methodological landscape, there exist several quite different approaches for incorporating risk in both the outranking and metric methods. Future research might examine:

1. how similar the results of the different treatments of risk within each methodology are,
2. qualitative issues such as the ease of use of the various methodologies and the scope provided for learning about the decision problem,
3. whether any methodologies are specifically suited to certain decision contexts.

At least some of these questions could be investigated using the simulation frameworks proposed for the metric and outranking methods in sections 4.5 and 4.6 respectively. It could be hoped that the results of such research would provide the input to further research providing an integrated and unified treatment of risk in each of the outranking and metric methods.

Development of Integrated Scenario-based MCDM Methods

In the course of this thesis we have laid out a framework for using a scenario-based methodology as a treatment of uncertainty in MCDM. For the meaningful integration of two schools of thought such as MCDM and scenario planning, it is important to consider potentially problematic areas where the two schools have different and possibly conflicting opinions, as well as specific instances to which the integrated methodology is particularly well-suited, and which can serve as an impetus to further integration. With this in mind, future research might consider the following areas.

1. An important question relates to the nature of the aggregation over scenarios, on the basis that scenario planning and MCDM hold quite different views as to what types of aggregation are appropriate. Future research might consider investigating different notions of robustness over scenarios in order to allow more flexibility in the decision making process, as well as the feasibility of using different forms of aggregation in order to obtain different views on the decision problem.
2. Within the outranking methodology, future research could consider the feasibility of simplifying the methodologically attractive but conceptually difficult distributive outranking method of d'Avignon and Vincke [19] by making use of scenario planning. Such research might take the form of either some initial practical case studies or a simulation approach similar to that outlined in chapter 4.6.

3. In a similar fashion to the outranking methods, research relating to metric methods could be directed into investigating the feasibility of using scenario planning as a basis for extending the recourse ideology to the case of goal programming under conditions of risk. Research should be directed into the methodological and algorithmic feasibility of such a model, with particular emphasis on the potential for modelling different behavioural aspects of the decision maker by employing different aggregation metrics within and between scenarios.
4. The issue of changing preference structures over scenarios also could be addressed by future research. The essential questions to be considered are whether the qualitative information that is made available during the construction of the scenarios leads the decision maker to change his or her preferences dependent on scenario, and whether this change affects the results of the analysis to a sufficient extent to warrant being included in the analysis. A related practical question is whether a decision maker is able and willing to assume the increased cognitive burden of considering preferences within each scenario.

Scenario Selection Policies

As we have mentioned, we believe that the issue of diversity is central to further integration of scenario planning and MCDM. The construction of a diverse set of scenarios is essentially a problem structuring problem. However, without a good idea as to what constitutes 'diversity', progress in developing structural tools is unlikely. Future research should therefore first be directed into:

1. providing operationally useful definitions of diversity as it is understood by those employing scenario planning techniques i.e. in a descriptive sense,
2. providing normative definitions that might guide the selection of diverse scenarios while avoiding the deterioration of results experienced in the simulations presented here.
3. simulating a broader range of selection policies than the ones used here. In particular, the possibility of using artificially constructed 'quartile' scenarios representing 'good', 'average', and 'bad' performance should be addressed.
4. simulating some of the claimed advantages of using scenario planning in a quantitative fashion i.e. illuminating some previously unseen scenarios or alternatives, learning benefits.

The fruits of such research might in turn suggest new problem structuring approaches for the construction of such scenarios. In all cases, it is important to bear in mind the concerns of both MCDM and scenario planning philosophies.

6.4 A Final Comment

The multicriteria decision making methodologies provide practical frameworks for the consideration of those difficult problems relating to the intricacies and idiosyncracies of human choice. The scope of the problem is such that an all-encompassing ‘solution’ is a fallacy and a diversion from the principal condition that an MCDM framework should remain usable by a diligent decision maker. This seems particularly important in the complex environments in which uncertainty is prevalent. Although several useful models exist, perhaps more practically meaningful methodological developments might only be achieved through a persistent effort to apply MCDM to an area in which uncertainty is an integral and unavoidable aspect, for example portfolio selection. In any decision problem, practitioners and theoreticians alike would be well served by the advice ‘to look for precision in each class of things just as far as the nature of the subject admits’ [91].

University of Cape Town

Bibliography

- [1] M. Allais. Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'école Américaine. *Econometrica*, 21:503–546, 1953.
- [2] A. Arbel and Y. Orgler. An application of the AHP to bank strategic planning: The mergers and acquisitions process. *European Journal of Operational Research*, 48:27–37, 1990.
- [3] S. Azondékon and J. Martel. Value of additional information in multicriterion analysis under uncertainty. *European Journal of Operational Research*, 117:45–62, 1999.
- [4] S. Baas and H. Kwakernaak. Rating and ranking of multiple-aspect alternatives using fuzzy sets. *Automatica*, 13(1):47–58, 1977.
- [5] E. Ballestero. A compromise approach to specification and optimization. *Journal of Multi-criteria Decision Analysis*, 6:11–16, 1997.
- [6] E. Ballestero. Stochastic goal programming: A mean-variance approach. *European Journal of Operational Research*, 131:476–481, 2001.
- [7] V. Belton and T. Stewart. *Multiple Criteria Decision Analysis: An Integrated Approach*. Kluwer Academic Publishers, Boston, 2002.
- [8] P. Bonissone. *The problem of linguistic approximation in systems analysis*. PhD thesis, UCB Berkeley, 1979.
- [9] D. Bouyssou. Outranking relations: do they have special properties? *Journal of Multi-criteria Decision Analysis*, 5:99–111, 1996.
- [10] J. Brans, P. Vincke, and B. Mareschal. How to select and how to rank projects: The PROMETHEE method. *European Journal of Operational Research*, 24:228–238, 1986.
- [11] D. Bunn and A. Salo. Forecasting with scenarios. *European Journal of Operational Research*, 68:291–303, 1993.
- [12] B. Cadet. Contextual Effects and the Influence of the Frame on Decision Making. In Munier and Machina [63], pages 17–34.

- [13] E. Carbone and J. Hey. Estimation of expected utility and non-expected utility preference functionals using complete ranking data. In Munier and Machina [63], pages 119–140.
- [14] N.-B. Chang and S. Wang. A fuzzy goal programming approach for the optimal planning of metropolitan solid waste management systems. *European Journal of Operational Research*, 99:303–321, 1997.
- [15] A. Charnes and W. Cooper. Optimal estimation of executive compensation by linear programming. *Management Science*, 1:138–151, 1955.
- [16] S.-J. Chen and C.-L. Hwang. *Fuzzy Multiple Attribute Decision Making: Methods and Applications*. Springer-Verlag, Berlin, 1992.
- [17] B. Contini. A stochastic approach to goal programming. *Operations Research*, 3:576–586, 1968.
- [18] K. Darby-Dowman, S. Barker, E. Audsley, and D. Parsons. A two-stage stochastic programming with recourse model for determining robust planting plans in horticulture. *Journal of the Operational Research Society*, 51:83–89, 2000.
- [19] G. d’Avignon and P. Vincke. An outranking method under uncertainty. *European Journal of Operational Research*, 36:311–321, 1988.
- [20] P. De, D. Acharya, and K. Sahu. A Chance-Constrained Goal Programming Model for Capital Budgeting. *Journal of the Operational Research Society*, 33:635–638, 1982.
- [21] B. Dendrou, S. Dendrou, and E. Houstis. Multiobjective decision analysis for engineering systems. *Computers and Operations Research*, 7:301–312, 1980.
- [22] A. Dimitras, R. Slowinski, R. Susmaga, and C. Zopoundis. Business failure prediction using rough sets. *European Journal of Operational Research*, 114:263–280, 1999.
- [23] R. Donckels. Regional multiobjective planning under uncertainty: A stochastic goal programming formulation. *Journal of Regional Science*, 17(2):207–216, 1977.
- [24] D. Dubois and H. Prade. Putting rough sets and fuzzy sets together. In Slowinski [81], pages 203–233.
- [25] D. Ellsberg. Risk, Ambiguity, and the Savage Axioms. *Quarterly Journal of Economics*, 75:643–669, 1961.
- [26] P. Fishburn. Foundations of Risk Measurement. I. Risk as probable loss. *Management Science*, 30(4):396–405, 1984.

- [27] S. French. *Decision Theory: An Introduction to the Mathematics of Rationality*. Ellis Horwood, Chichester, 1988.
- [28] S. French. Uncertainty and imprecision: Modelling and analysis. *Journal of the Operational Research Society*, 46:70–79, 1995.
- [29] T. Gal, T. Stewart, and T. Hanne, editors. *Multicriteria Decision Making: Advances in MCDM Models, Algorithms, Theory, and Applications*. Kluwer Academic Publishers, Dordrecht, 1999.
- [30] P. Goodwin and G. Wright. *Decision Analysis for Management Judgement*. John Wiley & Sons, Chichester, 2nd edition, 1998.
- [31] M. Goumas and V. Lygerou. An extension of the PROMETHEE method for decision making in a fuzzy environment: Ranking of alternative energy exploration projects. *European Journal of Operational Research*, 123:606–613, 2000.
- [32] S. Greco, B. Matarazzo, and R. Slowinski. Rough approximation of a preference relation by dominance relations. *European Journal of Operational Research*, 117:63–83, 1999.
- [33] S. Greco, B. Matarazzo, and R. Slowinski. Rough sets theory for multicriteria decision analysis. *European Journal of Operational Research*, 129:1–47, 2001.
- [34] W. Hallerbach and J. Spronk. A multi-dimensional framework for portfolio management. In M. Karwan, J. Spronk, and J. Wallenius, editors, *Essays in decision making: A volume in honour of Stanley Zionts*. Springer-Verlag, Berlin, 1997.
- [35] E. Hannan. Linear programming with multiple fuzzy goals. *Fuzzy Sets and Systems*, 6:235–248, 1981.
- [36] C. Huang, D. Kira, and I. Vertinsky. Stochastic dominance rules for multiattribute utility functions. *Review of Economic Studies*, 41:611–615, 1978.
- [37] J. Huff, A. Huff, and H. Thomas. Strategic renewal and the interaction of cumulative stress and inertia. *Strategic Management Journal*, 13:55–75, 1992.
- [38] E. Jacquet-Lagrèze. Modelling preferences among distributions using fuzzy relations. In H. Jungermann and G. De Zeeuw, editors, *Models and experiments in risk and rationality*, pages 99–114. D. Reidel Publishers, Dordrecht, 1977.
- [39] E. Jacquet-Lagrèze and Y. Siskos. Preference disaggregation: 20 years of MCDA experience. *European Journal of Operational Research*, 130:233–245, 2001.
- [40] J. Jia and J. Dyer. A standard measure of risk and risk-value models. *Management Science*, 42(12):1691–1705, 1996.

- [41] D. Kahneman and A. Tversky. Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–291, 1979.
- [42] P. Kall and S. Wallace. *Stochastic Programming*. John Wiley & Sons, Chichester, 1994.
- [43] T. Kalu. Capital budgeting under uncertainty: An extended goal programming approach. *International Journal of Production Economics*, 58:235–251, 1999.
- [44] A. Kandel. *Fuzzy mathematical techniques with applications*. Addison-Wesley, Massachusetts, 1986.
- [45] A. Kanudia and P. Shukla. Modelling of uncertainties and price elastic demands in energy-environment planning for india. *Omega*, 26(3):409–423, 1998.
- [46] D. Keefer and S. Bodily. Three-point approximations for continuous random variables. *Management Science*, 29(5):595–609, 1983.
- [47] R. Keeney. *Value-Focussed Thinking*. Harvard University Press, Cambridge, 1992.
- [48] R. Keeney and H. Raiffa. *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*. John Wiley & Sons, New York, 1976.
- [49] A. Keown and B. Taylor. A Chance-Constrained Integer Goal Programming Model for Capital Budgeting in the Production Area. *Journal of the Operational Research Society*, 31:579–589, 1980.
- [50] G. Klir and T. Folger. *Fuzzy Sets, Uncertainty, and Information*. Prentice Hall, New Jersey, 1988.
- [51] A. Korhonen. Strategic financial management in a multinational financial conglomerate: A multiple goal stochastic programming approach. *European Journal of Operational Research*, 128:418–434, 2001.
- [52] H. Kunreuther, J. Linnerooth, and J. Vaupel. A decision-process perspective on risk and policy analysis. *Management Science*, 30(4):475–485, 1984.
- [53] R. Lahdelma and P. Salminen. Pseudo-criteria versus linear utility functions in stochastic multi-criteria acceptability analysis. *European Journal of Operational Research*, 141:454–469, 2002.
- [54] S. Lee and D. Olson. Goal programming. In Gal et al. [29], chapter 8.
- [55] D. Lehman. Generalized qualitative probability: Savage revisited. In *Proceedings of the Conference on Uncertainty in AI*, pages 381–388. 1996.

- [56] B. Liu and K. Iwamura. Modelling stochastic decision systems using dependent-chance programming. *European Journal of Operational Research*, 101:193–203, 1997.
- [57] F. Lootsma, P. Boonekamp, R. Cooke, and F. Van Oostvoorn. Choice of a long-term strategy for the national electricity supply via scenario analysis and multi-criteria analysis. *European Journal of Operational Research*, 48:189–203, 1990.
- [58] M. Machina. Review of ‘Generalized Expected Utility Theory: The Rank-Dependent Model’. *Journal of Economic Literature*, 32(3):1237–1238, 1994.
- [59] H. Markowitz. Portfolio selection. *Journal of Finance*, 7(1):77–91, 1952.
- [60] J. Martel, S. Azondékon, and K. Zaras. Preference relations in multicriterion analysis under risk. *Belgian Journal of Operations Research, Statistics and Computer Science*, 31(3):55–83, 1992.
- [61] J. Martel and G. d’Avignon. Projects ordering with multicriteria analysis. *European Journal of Operational Research*, 10:56–69, 1982.
- [62] J. Martel, G. d’Avignon, and G. Couillard. A fuzzy outranking relation in multicriteria decision making. *European Journal of Operational Research*, 25:258–271, 1986.
- [63] B. Munier and M. Machina, editors. *Models and experiments in risk and rationality*. Kluwer Academic Publishers, Dordrecht, 1994.
- [64] R. Narasimhan. Goal programming in a fuzzy environment. *Decision Sciences*, 11:325–336, 1980.
- [65] B. Naslund. A model of capital budgeting under risk. *Journal of Business*, 39:257–271, 1966.
- [66] P. Odom, R. Shannon, and B. Buckles. Multi-goal subset selection problems under uncertainty. *AIIE Transactions*, 11(1):61–69, 1979.
- [67] Z. Pawlak. Rough sets. *International Journal of Information and Computer Sciences*, 11:341–356, 1982.
- [68] Z. Pawlak. *Rough Sets: Theoretical Aspects of Reasoning about Data*. Kluwer Academic Publishers, Dordrecht, 1991.
- [69] Z. Pawlak. Rough set approach to knowledge-based decision support. *European Journal of Operational Research*, 99:48–57, 1997.
- [70] Z. Pawlak and R. Slowinski. Rough set approach to multi-attribute decision analysis. *European Journal of Operational Research*, 72:443–459, 1994.

- [71] J.-C. Pomerol. Scenario development and practical decision making under uncertainty. *Decision Support Systems*, 31:197–204, 2001.
- [72] A. Prékopa. *Stochastic Programming*. Kluwer Academic Publishers, London, 1995.
- [73] J. Quiggin. *Generalized Expected Utility Theory*. Kluwer Academic Publishers, Dordrecht, 1993.
- [74] M. Rothschild and J. Stiglitz. Increasing risk: I. A definition. *Journal of Economic Theory*, 2:225–243, 1970.
- [75] B. Roy. *Methodologie Multicritere d'Aide à la Decision*. Economica, Paris, 1985.
- [76] B. Roy. A missing link in OR-DA: robustness analysis. *Foundations of Computing and Decision Sciences*, 23:141–160, 1998.
- [77] B. Roy and D. van der Pooten. The European school of MCDA: Emergence, Basic features and Current Works. *Journal of Multi-criteria Decision Analysis*, 5(1):22–38, 1996.
- [78] T. Saaty. *The Analytic Hierarchy Process*. McGraw-Hill, New York, 1980.
- [79] R. Sarin and M. Weber. Risk-value models. *European Journal of Operational Research*, 70:135–149, 1993.
- [80] H. Simon. *Administrative Behaviour*. The Free Press, New York, 1976.
- [81] R. Slowinski, editor. *Intelligent Decision Support: Handbook of Applications and Advances of the Rough Sets Theory*. Kluwer Academic Publishers, Dordrecht, 1992.
- [82] R. Slowinski, C. Zopoundis, and A. Dimitras. Prediction of company acquisition in Greece by means of the rough set approach. *European Journal of Operational Research*, 100:1–15, 1997.
- [83] C. Starmer. Developments in non-expected utility theory: The hunt for a descriptive theory of choice under risk. *Journal of Economic Literature*, 38:332–382, 2000.
- [84] R. Steuer. *Multiple Criteria Optimization: Theory, Computation, and Application*. John Wiley & Sons, New York, 1986.
- [85] T. Stewart. Simplified approaches for multicriteria decision making under uncertainty. *Journal of Multi-criteria Decision Analysis*, 4:246–258, 1995.
- [86] T. Stewart. Robustness of additive value function methods in MCDM. *Journal of Multi-criteria Decision Analysis*, 5:301–309, 1996.

- [87] T. Stewart. Scenario analysis and multicriteria decision making. In J. Climaco, editor, *Multicriteria Analysis*, pages 519–528. Springer-Verlag, Berlin, 1997.
- [88] T. Stewart. Measurements of risk in fisheries management. *ORiON*, 14(1/2):1–15, 1998.
- [89] T. Stewart. Concepts of interactive programming. In Gal et al. [29], chapter 10.
- [90] T. Stewart. Evaluation and refinement of aspiration-based methods in MCDM. *European Journal of Operational Research*, 113:643–652, 1999.
- [91] E. Sylla. Political, Moral and Economic decisions and the origins of mathematical probability: The case of Jacob Bernoulli's 'The Art of Conjecturing'. In Von Furstenberg [99], chapter 2.
- [92] M. Tamiz, D. Jones, and C. Romero. Goal programming for decision making: An overview of the current state-of-the-art. *European Journal of Operational Research*, 111:569–581, 1998.
- [93] E. Turban. *Decision support and expert systems: Management support systems*. Macmillan, New York, 1988.
- [94] A. Tversky and D. Kahneman. Judgment under uncertainty: heuristics and biases. *Science*, 185:1124–1131, 1974.
- [95] A. Tversky and D. Kahneman. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5(4):297–323, 1992.
- [96] K. Van der Heijden. *Scenarios: The Art of Strategic Conversation*. John Wiley & Sons, New York, 1996.
- [97] P. Vincke. Analysis of multicriteria decision aid in Europe. *European Journal of Operational Research*, 25:160–168, 1986.
- [98] P. Vincke. Outranking approach. In Gal et al. [29], chapter 11.
- [99] G. Von Furstenberg, editor. *Acting under uncertainty: Multidisciplinary conceptions*. Kluwer Academic Publishers, Dordrecht, 1990.
- [100] P. Wack. Scenarios: shooting the rapids. *Harvard Business Review*, Nov/Dec:139–150, 1985.
- [101] P. Wack. Scenarios: uncharted water ahead. *Harvard Business Review*, Sept/Oct:73–89, 1985.
- [102] D. Watkins, D. McKinney, L. Lasdon, S. Nielsen, and Q. Martin. A scenario-based stochastic programming model for water supplies from the highland lakes. *International Transactions in Operational Research*, 7:211–230, 2000.

- [103] E. Weber and W. Bottom. An empirical evaluation of the transitivity, monotonicity, accounting and conjoint axioms for perceived risk. *Organizational Behaviour and Human Decision Processes*, 45:253–275, 1990.
- [104] A. Wiercbicki. Reference point approaches. In Gal et al. [29], chapter 9.
- [105] G. Winckler. Necessity, chance, and freedom. In Von Furstenberg [99], chapter 6.
- [106] G. Wright and P. Goodwin. Future-focussed thinking: Combining scenario planning with decision analysis. *Journal of Multi-criteria Decision Analysis*, 8:311–321, 1999.
- [107] G. Wu and R. Gonzalez. Nonlinear decision weights in choice under uncertainty. *Management Science*, 45(1):74–85, 1999.
- [108] C.-H. Yeh, H. Deng, and H. Pan. Multicriteria analysis for dredger dispatching under uncertainty. *Journal of the Operational Research Society*, 50:35–43, 1999.
- [109] M. Yilmaz. An information-expectation framework for decisions under uncertainty. *Journal of Multi-criteria Decision Analysis*, 1:65–80, 1992.
- [110] L. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.
- [111] K. Zaras. Rough approximation of a preference relation by a multi-attribute stochastic dominance for determinist and stochastic evaluation problems. *European Journal of Operational Research*, 130:305–314, 2001.
- [112] K. Zaras and J. Martel. Multiattribute analysis based on stochastic dominance. In Munier and Machina [63], pages 225–248.
- [113] H.-J. Zimmerman. Fuzzy programming and linear programming with several objective functions. *Fuzzy Sets and Systems*, 1:45–55, 1978.
- [114] H.-J. Zimmerman. *Fuzzy Sets, Decision Making, and Expert Systems*. Kluwer Academic Publishers, Boston, 1987.
- [115] H.-J. Zimmerman. An application-oriented view of modeling uncertainty. *European Journal of Operational Research*, 122:190–198, 2000.