

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.

# Simplified models for multi-criteria decision analysis under uncertainty

Ian N. Durbach



Thesis presented for the degree of Doctor of Philosophy in the Department of Statistical Sciences at the University of Cape Town.

Supervised by Professor Theodor Stewart

October 2010

## **Acknowledgements**

I would like to offer my sincere thanks to the following people and organisations for their assistance and support along the way: my supervisor Prof. Theo Stewart; colleagues in the Department of Statistical Sciences; Simon French and the Decision Sciences Research Center at Manchester Business School, where the work on which Chapter 4 is based was carried out; those working at the Energy Research Center, especially Stephen Davis and Brett Cohen; Sustainable Energy Africa; Graham Barr; Jan Hofmeyr and Synovate; Butch Rice and Research Surveys; the National Research Foundation, which provided funding for parts of the project; and friends and family.

## Abstract

When facilitating decisions in which some performance evaluations are uncertain, a decision must be taken about how this uncertainty is to be modelled. This involves, in part, choosing an uncertainty format – a way of representing the possible outcomes that may occur. It seems reasonable to suggest – and is an aim of the thesis to show – that the choice of how uncertain quantities are represented will exert some influence over the decision-making process and the final decision taken.

Many models exist for multi-criteria decision analysis (MCDA) under conditions of uncertainty; perhaps the most well-known are those based on multi-attribute utility theory [MAUT, e.g. 147], which uses probability distributions to represent uncertainty. The great strength of MAUT is its axiomatic foundation, but even in its simplest form its practical implementation is formidable, and although there are several practical applications of MAUT reported in the literature [e.g. 39, 270] the number is small relative to its theoretical standing. Practical applications often use simpler decision models to aid decision making under uncertainty, based on uncertainty formats that ‘simplify’ the full probability distributions (e.g. using expected values, variances, quantiles, etc). The aim of this thesis is to identify decision models associated with these ‘simplified’ uncertainty formats and to evaluate the potential usefulness of these models as decision aids for problems involving uncertainty. It is hoped that doing so provides some guidance to practitioners about the types of models that may be used for uncertain decision making.

The performance of simplified models is evaluated using three distinct methodological approaches – computer simulation, ‘laboratory’ choice experiments, and real-world applications of decision analysis – in the hope of providing an integrated assessment. Chapter 3 generates a number of hypothetical decision problems by simulation, and within each problem simulates the hypothetical application of MAUT and various simplified decision models. The findings allow one to assess how the simplification of MAUT models might impact results, but do not provide any general conclusions because they are based on hypothetical decision problems and cannot evaluate practical issues like ease-of-use or the ability to generate insight that are critical to good decision aid. Chapter 4 addresses some of these limitations by reporting an experimental study consisting of choice tasks presented to numerate but unfacilitated participants. Tasks involved subjects selecting one from a set of five alternatives with uncertain attribute evaluations, with the format used to represent uncertainty and the number of objectives for the choice varied as part of the experimental design. The study is limited by the focus on descriptive rather than real prescriptive decision making, but has implications for prescriptive decision making

practice in that natural tendencies are identified which may need to be overcome in the course of a prescriptive analysis.

Chapter 5 complements the experimental studies by reporting three real-world applications using simplified decision models. The first application involves the construction of a framework for evaluating the effectiveness of intervention programs aimed at reducing household electricity consumption, and uses quantiles to represent uncertainty. The second constructs a risk-value framework for depicting different slot machine varieties on the floor of a large casino in the Western Cape, and uses probabilities of poor performance. In the third application, a group of market researchers must evaluate how one “client” alternative performs in relation to its competitors on the basis of multi-attribute evaluations collected in a survey; expected values are used to represent uncertain evaluations. The applications are useful in addressing more practical issues around the use of simplified models.

The most important findings drawn collectively from the three methodological approaches are:

1. The use of any of the simplified uncertainty formats considered in this thesis rather than probability distributions can be justified.
2. A model using quantiles is the most promising of the simplified models and should be used as a ‘default’ option in the absence of other information.
3. A model using variances is the least promising of the simplified models and should not be used unless there are compelling reasons to do so.
4. Scenarios result in relatively poor approximations of MAUT and are primarily devices for helping decision makers to gain a better understanding of the causes of uncertainty and their consequences.
5. Placing elements of simplified uncertainty formats (scenarios, quantiles, variances and other explicit risk attributes) in the second level of the objectives hierarchy can offer a useful uncertainty-orientated view of a decision problem.
6. Deviations from linearity in marginal utility functions increase the attractiveness of MAUT relative to all of the simplified models.
7. Increases in the size of a decision problem (number of alternatives and attributes) increase the attractiveness of all of the simplified models relative to MAUT.
8. Errors in the assessment of inputs to all simplified models cause substantial deteriorations in model accuracy.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1-1</b>
1.1	Background . . . . .	1-1
1.2	Statement of the problem . . . . .	1-1
1.3	Methodology . . . . .	1-2
1.4	Objectives . . . . .	1-4
1.5	Limitations . . . . .	1-5
1.6	Plan of development . . . . .	1-5
<b>2</b>	<b>Literature review: uncertainty in decision analysis</b>	<b>2-1</b>
2.1	Methods for eliciting assessments of uncertainty . . . . .	2-1
2.1.1	Uncertainty formats . . . . .	2-1
2.1.2	Assessment and framing effects . . . . .	2-3
2.2	Decision analysis with probability distributions . . . . .	2-8
2.2.1	Multi-attribute utility theory . . . . .	2-8
2.2.2	Pairwise comparisons of probability distributions . . . . .	2-9
2.2.3	Models simulating from probability distributions . . . . .	2-11
2.3	Decision analysis with decision weights . . . . .	2-12
2.3.1	Empirical evidence . . . . .	2-12
2.3.2	Descriptive theories of choice under uncertainty . . . . .	2-14
2.3.3	Developments in prescriptive multi-criteria decision aid . . . . .	2-15
2.4	Decision analysis with explicit risk attributes . . . . .	2-18
2.4.1	The measurement of risk . . . . .	2-18
2.4.2	Multi-attribute applications . . . . .	2-20
2.4.3	Models using quantiles . . . . .	2-20
2.5	Decision analysis with fuzzy sets . . . . .	2-23
2.5.1	Aggregation models for fuzzy evaluation . . . . .	2-24
2.5.2	Methods for ranking fuzzy numbers . . . . .	2-27
2.6	Decision analysis with scenarios . . . . .	2-29
2.6.1	Scenarios in scenario planning . . . . .	2-30
2.6.2	Scenarios in multi-criteria analysis . . . . .	2-31

<b>3</b>	<b>A simulation study investigating the effect of uncertainty representation on decision making</b>	<b>3-1</b>
3.1	Introduction . . . . .	3-1
3.2	Summary of simplified decision models used . . . . .	3-2
3.2.1	Models using expected values and explicit risk attributes . . . . .	3-2
3.2.2	Models using quantiles . . . . .	3-2
3.2.3	Models using fuzzy numbers . . . . .	3-3
3.2.4	Models using scenarios . . . . .	3-3
3.3	Research aims and hypotheses . . . . .	3-4
3.4	Design of the simulation experiment . . . . .	3-5
3.4.1	Generating realizations from $Z_{ij}$ . . . . .	3-5
3.4.2	Generating inputs to the simplified models . . . . .	3-7
3.4.3	Generating errors in the assessment of uncertainty information . . . . .	3-9
3.4.4	Generating preference structures . . . . .	3-9
3.4.5	Simulating the application of an additive MAUT model . . . . .	3-9
3.4.6	Simulating the application of the simplified models . . . . .	3-9
3.4.7	Comparing results of the MAUT and simplified models . . . . .	3-11
3.4.8	Parameter values used in the simulations . . . . .	3-11
3.5	Results . . . . .	3-13
3.6	Discussion . . . . .	3-17
<b>4</b>	<b>A choice experiment investigating the effect of uncertainty representation on decision making</b>	<b>4-1</b>
4.1	Introduction . . . . .	4-1
4.2	Research hypotheses . . . . .	4-2
4.3	Details of choice experiment . . . . .	4-3
4.3.1	Recruitment and sample selection . . . . .	4-3
4.3.2	Description of the task . . . . .	4-4
4.3.3	Verbal commentary and facilitator involvement . . . . .	4-4
4.3.4	Design of investment alternatives . . . . .	4-5
4.3.5	Uncertainty formats . . . . .	4-6
4.3.6	Experimental design . . . . .	4-7
4.3.7	Experimental outcomes . . . . .	4-8
4.3.8	Model building . . . . .	4-9
4.4	Results . . . . .	4-10
4.4.1	Factors affecting accuracy . . . . .	4-10
4.4.2	Factors affecting assessments of difficulty . . . . .	4-12
4.4.3	Factors affecting the type of alternative chosen . . . . .	4-15

4.4.4	Analysis of verbal commentaries . . . . .	4-16
4.5	Discussion . . . . .	4-20
4.5.1	Reconciling the simulation and choice experiments . . . . .	4-20
4.5.2	Practical issues . . . . .	4-22
<b>5</b>	<b>Applications of decision support with simplified representations of un-</b>	
	<b>certainty</b>	<b>5-1</b>
5.1	Introduction . . . . .	5-1
5.2	Stochastic multi-criteria acceptability analysis (SMAA) . . . . .	5-2
5.3	Integrating simplified uncertainty formats with SMAA . . . . .	5-5
5.3.1	Expected values . . . . .	5-5
5.3.2	Explicit risk attributes . . . . .	5-5
5.3.3	Quantiles . . . . .	5-6
5.3.4	Fuzzy numbers . . . . .	5-6
5.3.5	Scenarios . . . . .	5-7
5.4	Decision support for evaluating electricity saving options . . . . .	5-7
5.4.1	Background to problem . . . . .	5-7
5.4.2	Problem structuring . . . . .	5-8
5.4.3	Results . . . . .	5-13
5.4.4	Follow-up . . . . .	5-23
5.5	Decision support for evaluating slot machines . . . . .	5-23
5.5.1	Background to problem . . . . .	5-23
5.5.2	Problem structuring . . . . .	5-24
5.5.3	Results . . . . .	5-26
5.5.4	Follow-up . . . . .	5-29
5.6	Decision support for evaluating brands in consumer markets . . . . .	5-30
5.6.1	Background to problem . . . . .	5-30
5.6.2	Problem structuring . . . . .	5-31
5.6.3	Results . . . . .	5-35
5.6.4	Follow-up . . . . .	5-41
5.7	Conclusions . . . . .	5-41
<b>6</b>	<b>Conclusions</b>	<b>6-1</b>
6.1	Findings and recommendations . . . . .	6-1
6.2	Questions for future research . . . . .	6-6
<b>A</b>	<b>Supplementary material for Chapter 3</b>	<b>A-1</b>
A.1	Equivalence of quantile and fuzzy decision models . . . . .	A-1



<b>B</b>	<b>Supplementary material for Chapter 4</b>	<b>B-1</b>
B.1	Documents given to study participants . . . . .	B-1
B.2	Decision models used . . . . .	B-19
B.2.1	Additive MAUT model . . . . .	B-19
B.2.2	Expected values . . . . .	B-19
B.2.3	Expected values and standard deviations . . . . .	B-19
B.2.4	Quantiles . . . . .	B-19
B.2.5	Three-point approximations/fuzzy numbers . . . . .	B-19
B.2.6	Scenarios . . . . .	B-20
B.3	Utility function parameters used . . . . .	B-20
B.4	Accuracy results for all utility functions . . . . .	B-20

University of Cape Town

# List of Figures

2.1	A value tree illustrating a hypothetical MCDA process where uncertain evaluations are represented using quantiles. . . . .	2-21
2.2	An alternate value tree formulation for quantile-based MCDA. In this tree, quantiles are represented in the second level of the objectives hierarchy rather than as lower-level criteria. . . . .	2-22
3.1	Outline of a single simulation run . . . . .	3-6
3.2	Mean utility loss experienced by each simplified model. Unshaded and (three) shaded points show the average utility loss in error-free simulations and with assessment errors of 10%, 20%, and 30%. . . . .	3-14
4.1	Parameter estimates for models of decision accuracy. The effects of uncertainty formats are shown in (a), other effects are shown in (b). . . . .	4-11
4.2	Parameter estimates for models of perceived difficulty and confidence. The effects of uncertainty formats are shown in (a), other effects are shown in (b).	4-13
4.3	Modes of uncertainty format evaluated according to their accuracy and difficulty in multi-objective problems. In (a) accuracy and difficulty are predicted values obtained from the fitted models; in (b) they are obtained by holistic judgements made by subjects after the set of tasks was completed.	4-14
4.4	Parameter estimates for the rank of the chosen alternative according to specific characteristics: (a) expected values and maxima, and (b) standard deviations and minima. . . . .	4-16
5.1	Extract from spreadsheet used by decision makers to inform selection of electricity saving options. . . . .	5-11
5.2	Value tree for selecting promising electricity savings options. . . . .	5-13
5.3	Acceptability indices for electricity savings options obtained with an adapted SMAA-O model. . . . .	5-17
5.4	Central weight vectors for electricity savings options obtained with an adapted SMAA-O model. All figures are based upon Model 3. . . . .	5-21
5.5	Illustration of three typical ‘winning lines’ on a slot machine. . . . .	5-24

5.6	Results of a SMAA-2 analysis showing, (a) acceptability indices for the first four ranks for all machines having any of the corresponding acceptability indices greater than zero, (b) central importance weights for all machines with rank-one acceptability greater than zero. . . . .	5-28
A.1	Illustration of computations performed by fuzzy decision model . . . . .	A-1
B.1	Estimates of the effects of uncertainty format on accuracy using different utility functions (Sets 1 to 6). . . . .	B-21
B.1	Estimates of the effects of uncertainty format on accuracy using different utility functions (Sets 7 to 12). . . . .	B-22

University of Cape Town

# List of Tables

2.1	A guide to recent fuzzy ranking methods . . . . .	2-28
3.1	Parameter values used to simulate hypothetical decision problems . . . . .	3-12
3.2	Parameters used to simulate the application of simplified decision models . . . . .	3-12
3.3	Increases in mean utility loss caused by assessment errors. Results are only reported for simulations where dominated alternatives have been removed. Simplified models are represented by their best-performing versions, after excluding models using optimal risk or quantile weights. . . . .	3-16
3.4	Average utility losses at different levels of the simulated decision problem parameters. Averages are calculated over all levels of the remaining parameters. Results are only reported for simulations without dominated alternatives. The same models are used as in Table 3.3: the variance model uses Kirkwood weights; the probability of poor performance model uses a 0.50 cutoff and a risk multiplier of 0.25, both quantile models use equal quantile weights; the fuzzy model uses 5% and 95% quantiles with $\Theta = 0.5$ , and all scenario models use mean scenarios. . . . .	3-18
4.1	A classification of the experimental uncertainty formats according to their effect on naïve decision making . . . . .	4-24
5.1	Constructed scale used to measure the “ease of use” criterion in the evaluation of electricity saving options. . . . .	5-9
5.2	Decision table for evaluation of electricity saving options . . . . .	5-13
5.3	Attribute evaluations for five slot machines . . . . .	5-27
5.4	Lists of attributes used to evaluate alternatives in the two market research studies . . . . .	5-33
5.5	Associations between survey measures and overall market share for all brands with market share greater than 5% in Dataset 1 . . . . .	5-36
5.6	Associations between survey measures and overall market share for all brands with market share greater than 5% in Dataset 2 . . . . .	5-36
5.7	Associations between survey measures and share of purchases in Dataset 1 . . . . .	5-37
5.8	Associations between survey measures and share of purchases in Dataset 2 . . . . .	5-37

5.9 Associations between survey measures and defection behaviour in Dataset 1 5-38

5.10 Associations between survey measures and defection behaviour in Dataset 2 5-39

5.11 Associations between survey measures and share of purchases in Dataset 1 . 5-40

5.12 Associations between survey measures and share of purchases in Dataset 2 . 5-40

B.1 Parameter estimates for utility functions used to compute utility loss. Reference point values ( $\tau_j$ ) are given in thousands of pounds. . . . . B-20

University of Cape Town

# Chapter 1

## Introduction

### 1.1 Background

This thesis is based on three uncontroversial statements about decision making. First, in comparing alternative courses of action it is often useful to value their performances on a number of different attributes rather than attempt an overall valuation directly. Second, in nearly all cases there will be no one alternative that performs best on all attributes. It is partly for this reason that overall valuations are difficult to make directly. The comparison of different alternatives requires that the trade-offs between performance on different attributes be addressed and themselves subjected to some kind of valuation. It also requires that these (now weighted) evaluations be aggregated in some way to arrive at a final decision. Third, in many cases the evaluation of alternatives will be complicated by their performance on at least some attributes not being known with certainty. It is the treatment of cases in which all three of these statements are true – decisions involving multiple conflicting objectives and attribute evaluations that are in some way uncertain – that form the subject of the thesis.

### 1.2 Statement of the problem

When facilitating decisions in which performance evaluations are uncertain, a decision must be taken about how this uncertainty is to be modelled. This involves, in part, choosing an uncertainty format – a way of representing the possible outcomes that may occur. At one extreme level of detail, one may write out all the outcomes that are possible for a given alternative and associate a probability with each possible outcome. At the opposite extreme, one may use a single (probably location) measure such as the mean, median, or mode. Between these extremes lie a host of possible compromises: measures of spread such as the variance or range; sets of quantiles; probabilities of achieving specified performance levels; and piecewise approximations of the full probability distributions, to name a few. A related step in modelling decisions with uncertain outcomes involves choosing a model for representing preferences over uncertain outcomes. This choice is closely related to the

choice of an uncertainty format, which may suggest or even impose a particular preference model (and *vice versa*). Throughout this thesis, any uncertainty format summarising the full probability distribution is referred to as a ‘simplified uncertainty format’, and any decision model using a simplified uncertainty format is termed a ‘simplified decision model’.

It seems reasonable to suggest – and is an aim of the thesis to show – that the choice of how uncertain quantities are represented will exert some influence over the decision-making process and the final decision taken. Decision makers may assess some uncertainty formats more accurately than others, which may give decision models based on some pieces of information an advantage over others that are poorly assessed. Even if correctly assessed most uncertainty formats summarise the full range of possible outcomes, which may bias decisions in favour of those alternatives that perform relatively better on whatever summary measures happen to be used. Certain aspects of decision making (e.g. making trade-offs between performances on different objectives, or between risk and return on a single objective; or eliminating dominated alternatives) may also be easier with (or more strongly suggested by) some uncertainty formats than others.

Of the many models that exist for multi-criteria decision analysis (MCDA) under conditions of uncertainty, perhaps the most well-known are those based on multi-attribute utility theory [MAUT, e.g. 147]. The great strength of MAUT is its axiomatic foundation “justifying the prescriptive approach provided the problem owners accept the related rationality assumptions” [116], but even in its simplest form the practical implementation of MAUT is formidable, requiring the assessment of full probability distributions over each attribute as well as trade-offs involving single- and multi-attribute lotteries. Although there are several practical applications of MAUT reported in the literature [e.g. 39, 270], this number is small relative to its theoretical standing. The view taken here is that practitioners often prefer to use simpler decision models to aid decision making under uncertainty. The aim of this thesis is to identify a number of such ‘simplified’ decision models that are currently in use and to evaluate the potential usefulness of these models as decision aids for problems involving uncertainty – that is, to establish the costs, if any, involved with the use of simplified uncertainty formats in decision analysis. In doing so I hope to provide some guidance to practitioners about the types of simplified models that are being used for uncertain decision making.

### 1.3 Methodology

Addressing the aim above requires that one specify the kinds of ‘costs’ which may arise as a result of using a simplified model and then establish theoretical or empirical results evaluating these costs. The precise outcomes used to evaluate the effectiveness of a decision

model are defined in the course of the thesis, but can be broadly characterised as measures of either “substantive” or “procedural” quality (in the usual terminology, substantive or procedural “rationality” [233]). A model or modelling process that is of a high substantive quality leads to a good outcome or course of action; a model or modelling process that is of a high procedural quality possesses certain properties that are assumed *a priori* to be desirable. Although procedural quality can give the appearance of being “amorphous, plastic, and somewhat arbitrary” [247], there is considerable agreement within the field of decision analysis about what properties good decision aid should possess. These include helping the decision maker to learn about or clarify the problem which faces them, as well as their own preferences; using concepts that are familiar to the decision maker, or can be clearly and unambiguously understood with reasonable effort; not requiring an inordinate amount of interaction time; yielding transparent and justifiable results; and challenging existing points of view and encouraging the exploration of potential courses of action [223, 117].

Evaluations of substantive and procedural quality require quite different methodological approaches. Evaluations of substantive quality have a long history of using computer simulation (e.g. [20, 243]), largely because doing so using real-world MCDA applications is difficult or impossible. This is because, prior to the decision process, preferences are not fully formed and merely awaiting elicitation. Rather these are at least partly formed by the process itself – the “constructivist” view of decision aid [31]. Identifying whether a selected alternative is ‘objectively’ good is thus difficult even for a single decision problem. Comparing different approaches would require that the same decision maker be facilitated through the independent construction of multiple simplified models, which seems impossible to achieve practically. In contrast the controlled but artificial environment of a simulation allows one to specify preferential information exactly and to vary certain aspects of the decision process independently of the others.

Evaluating procedural quality requires some form of human participation; simulation experiments cannot evaluate practical issues like a decision model’s ease-of-use or ability to generate insight into a problem. Real-world applications in the mode of “action research” [197] can be used. These provide clear results around practical issues but for MCDA – because of “constructivism” – any comparisons between different approaches can only be drawn qualitatively (though this can still be useful). The generality of any finding is also difficult to establish because the time and effort involved in a single application means that conclusions are usually based on only a small number of applications, and many elements within each application cannot be controlled for experimentally.

An intermediate approach which addresses some of the limitations of each of the method-



ologies above is to design simplified versions of a decision problem and present these to study participants in an experimental setting. This offers the researcher some degree of control while retaining the aspect of human decision making that makes real-world application so attractive. Although this might be a somewhat unconventional approach to research in MCDA, similar strategies have been regularly applied to the study of human decision making (e.g. [204, 142, 214, 233]). By designing the decision task carefully it is possible to obtain some indication of substantive quality by which to compare simulation results with human decision making. Elements of procedural quality can also be assessed by participants, although these will need to take into account the simplifications made to the real-world problem. By simplifying the decision problem it becomes possible to show each decision maker multiple problems, which allows for comparisons of different approaches within the same decision maker. In this thesis I refer to such experiments as “choice experiments”.

The approach followed in this thesis is to evaluate the performance of simplified models using each of these methodologies: computer simulation, choice experiments, and real-world applications. By integrating the results of all analyses I hope to provide a unified evaluation of the performance of simplified models for decision making under uncertainty.

## 1.4 Objectives

The broad aim of the thesis is to examine the way in which uncertain attribute evaluations are treated by models for multi-criteria decision analysis and in particular to provide an evaluation of the usefulness of models which in some way summarise the full distribution of possible outcomes. A more specific set of objectives is:

1. To describe currently available treatments of uncertainty in decision making and to place these into a coherent framework.
2. To use the controlled environment of a simulation experiment,
  - (a) to identify whether, and under what conditions, the simplified models are able to give a good approximation to the results that would be obtained in idealised conditions using MAUT.
  - (b) to assess the robustness of the simplified models to various environmental conditions.
3. To use an experimental setting involving a set of hypothetical decision tasks performed by human subjects
  - (a) to gain some idea of how unskilled decision makers go about making multi-criteria choices when outcomes are uncertain, and the quality of those decisions,

- (b) to investigate how the form in which uncertain outcomes are presented affects the quality and process of decision making,
  - (c) to investigate decision makers' preferences for different uncertainty formats.
4. To apply a small number of the simplified decision models to real-world decision problems in which uncertainty is a substantial component, and to evaluate their effectiveness.
  5. To construct a set of guidelines for modelling uncertain attribute evaluations in multi-criteria decisions.

## 1.5 Limitations

- The thesis considers only the type of uncertainty that relates to the physical randomness of future events, where the consequences of a particular choice are unknown because they depend in some way on future events. This is sometimes referred to as “external uncertainty” [246]. Many other types of uncertainty have been proposed – some examples are uncertainty about values and trade-off assessments, uncertainty about whether all criteria have been included, and uncertainty about the aim of the decision problem. These other uncertainties cannot be ignored entirely – they may be present in both the choice experiments of Chapter 4 and the real-world applications of Chapter 5 – but an integrated treatment of all types of uncertainties in multi-criteria modelling is beyond the scope of the thesis.
- With the exception of the literature review the thesis considers only value function methods. Other schools of MCDA (e.g. outranking, reference point methods) are excluded in the interest of keeping results concise. The axiomatic foundation of the value function approaches also make it easier to hypothesise an “idealised” decision process – other MCDA approaches would probably need to choose one particular variant or model as ‘ideal’ but could otherwise proceed along similar lines to those described in Chapters 3 and 4.
- The simulation and choice experiments conducted in Chapters 3 and 4 are based on the assumption that the aim of the analysis is to place the alternatives in a (not necessarily complete) rank order, that is the ranking “problematique” [222] is assumed in these chapters.

## 1.6 Plan of development

The following plan sketches the outline of the chapters that follow:

**Chapter 2** contains the literature review. Two broad streams of literature are reviewed.

The first covers the assessment of different uncertainty formats, and is drawn from the behavioural sciences – predominantly the psychology of measurement and perception. The second covers decision models employing different uncertainty formats: models using probabilities or probability-like quantities, quantiles, variances or variance-like quantities, fuzzy sets, and scenarios.

**Chapter 3** describes a simulation experiment used to analyse the performance of five simplified approaches for representing uncertainty in the context of multi-attribute utility/value theory: variances, probabilities of achieving some specified level of performance, quantiles, three-point approximations to the distributions (fuzzy sets), and scenarios. Performance is analysed for a variety of environmental conditions and decision maker characteristics. A number of implications for the remainder of the thesis are identified.

**Chapter 4** describes a second experiment investigating the effects of uncertainty formats on decision making. In contrast to the previous chapter, which uses a simulation environment, the focus of this chapter is on actual decision making behaviour, in the form of a choice experiment administered to 28 study participants. The same six uncertainty formats are used, and effects on decision making are tracked using four outcomes – the quality of the final choice, the characteristics of the alternative that is selected, the difficulty experienced in making a decision, and the process by which a decision was made. The results build on the simulation results by providing insights into how human decision makers make single- and multi-criteria choices in the presence of uncertainty (and some format for representing uncertainty) but in the absence of any facilitation.

**Chapter 5** describes three real-world applications involving decision analysis with simplified uncertainty formats. The first application involves the construction of a framework for evaluating the effectiveness of intervention programs aimed at reducing household electricity consumption, undertaken for an energy research group. The objective in this case was to provide a comparative evaluation of intervention projects. The second application constructs a risk-value framework for depicting different slot machine varieties and was undertaken for a large casino in the Western Cape. The main objective of this analysis is to describe the slot machines in a systematic and formalised way, allowing the casino to determine the types of players (in terms of attitudes towards playing) that they are catering for with their portfolio of machines. The third application provides decision support to a group of market researchers about how one “client” alternative performs in relation to its competitors on the basis of multi-attribute evaluations collected in a survey.

**Chapter 6** collects together the conclusions drawn in the previous chapters and states these as a number of findings and related recommendations for the practical support of decisions involving uncertain attribute evaluations. A final section concludes with some questions for future research emerging from the thesis.

University of Cape Town

## Chapter 2

# Literature review: uncertainty in decision analysis

At an intuitive level, uncertainty is easily recognised as a distinct lack of something – the absence of complete knowledge or certainty. However, just as there are many different kinds of knowledge, there are also many different types of uncertainty. The type of uncertainty considered in this thesis arises when the consequences of an action are unknown because they depend on future events. This is sometimes termed “external uncertainty” (e.g. [246]) because it relates to uncertainty about environmental conditions lying beyond the control of the decision maker. External uncertainty corresponds to uncertainty arising from physical randomness in French’s taxonomy of uncertainty [101], and to the inherent lack of information about the uncertain future in Zimmerman’s taxonomy [304]. In the remainder of the thesis I refer to external uncertainty simply as ‘uncertainty’.

## 2.1 Methods for eliciting assessments of uncertainty

### 2.1.1 Uncertainty formats

#### Probabilities

Consider a decision problem consisting of  $I$  alternatives denoted by  $a_i$ ,  $i \in \{1, \dots, I\}$ , evaluated on  $J$  attributes denoted by  $c_j$ ,  $j \in \{1, \dots, J\}$ . Let  $Z_{ij}$  be a random variable denoting the (stochastic) attribute evaluation of  $a_i$  on  $c_j$ , with an associated probability distribution governing its possible outcomes. A multivariate probability distribution governs the joint realisation of performance outcomes across all alternatives and all attributes. Independence conditions would dictate whether the full joint distribution could be constructed from the marginals or whether it would in principle have to be evaluated directly. This is ‘in principle’ because in all likelihood directly assessing the joint distribution would be impossibly complex. The use of probabilities and probability distributions is reviewed in Section 2.2.

## Decision weights

A large body of empirical research suggests that when making decisions people do not weigh probabilities linearly. That is, they transform probabilities of occurrence into what are known as “decision weights” [139, 290], and use these. This raises the question of *how* people transform probabilities. Research on this question is ongoing but many important results have been obtained, suggesting for example that people underweight large probabilities and overweight small probabilities, and also overweight low-ranked outcomes relative to higher-ranked outcomes. These findings as well as the general use of decision weights are reviewed as part of Section 2.3.

## Explicit risk measures

Explicit risk measures are generally attempts to summarise the full probability distribution using a single (or small number of) measures indicating how uncertain the performance of an alternative on a particular attribute is. The fundamental notion is that underlying uncertain attributes can be expressed in terms of both ‘value’ and ‘risk’ components. ‘Value’ components are typically based on expected values or some other central location measure. Specific risk measures are discussed in more detail in Section 2.4, but some common examples include the variance or standard deviation, inter-quartile or max-min ranges, specified quantiles of the distributions, the probability of obtaining a performance below some specified value, and coefficients of relative or absolute risk reduction.

## Fuzzy sets

Fuzzy set theory is a general theory for the modelling of imprecision. The fundamental notion in fuzzy set theory 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 inclusion.

## Scenarios

Scenarios are incomplete descriptions of how the future might unfold, with emphasis placed on the development of an internally-consistent chain of causal reasoning that allows the decision maker to gain understanding of the problem at hand and generate unusual insights into possible courses of action. The use of scenarios is more qualitative than any of the other measures discussed here, and form part of the ‘contextualised’ approaches (e.g. [136]) focusing on providing people with an informational context – in terms of descriptions of antecedents and consequences – in which to understand and interpret their risk. Although because of the incompleteness of their descriptions scenarios are different from ‘states

of the world' i.e. single realisations of the full multivariate distribution defined over all possible outcomes, it is often useful to think of them as representing groups of states that share similar important characteristics. The construction and use of scenarios in decision analysis is reviewed in Section 2.6.

### 2.1.2 Assessment and framing effects

The decision problem under conditions of uncertainty requires that the decision maker express his or her individual judgement about the uncertain quantities: which outcomes are possible, which outcomes are more likely than others, and so on. The processes whereby these expressions are obtained are referred to as 'assessment' or 'elicitation' procedures. Spetzler and Staël Von Holstein [238] define elicitation as "the process of extracting and quantifying individual judgement about uncertain quantities", while Garthwaite et al. [102] take the more explicitly constructivist view that it is "the process of formulating a person's knowledge and beliefs about one or more uncertain quantities into a (joint) probability distribution". Although this second definition reflects the overwhelming focus on *probability* elicitation, one could substitute any of the other forms of uncertainty representation. The other noteworthy feature is the implicit reference in both definitions to the role played by a facilitator or analyst who helps with the 'extracting' or 'formulating'.

### Probabilities and decision weights

Methods for eliciting probabilities – for example probability wheels, direct assessments, or other graphical aids – are well-established, and reviews can be found in [238] and chapter 10 of [265]. These elicitation methods aim to help the decision maker make "good" assessments, where this is usually interpreted (e.g. [102]) as assessments that are (a) internally consistent or coherent i.e. obeying certain logical conditions, in particular the laws of probability, (b) externally consistent or well-calibrated i.e. against any available data, and (c) self-consistent or reliable i.e. the assessments should be stable over repeated tests.

At around the same time that the elicitation methods came to prominence in decision analysis, a growing body of evidence suggested that peoples' assessments of probabilities often fell foul of all three of these requirements. Assessments were subject to systematic biases which, given certain conditions, could lead to substantial errors in judgements: internal-, external-, and self-inconsistency. This body of research is equally well-established and documented, and reviews can be found in [257, 138, 104, 105]. The "heuristics and biases" model proposes that people make use of three basic heuristics when assessing probabilities:

1. *Anchoring and adjustment* refers to the tendency to fix a probability (or other) judgment at an initial estimate, the "anchor", and then adjust it. Research suggests that the selection of an anchor may be influenced by irrelevant information, and that

the size of the adjustment away from the anchor is in many instances insufficient. Biases arising from the anchoring and adjustment heuristic include systematic over- or under-estimation and a tendency to assess conjunctive events as more likely than (equally probable) disjunctive ones.

2. *Availability* is the tendency to assess the frequency or probability of an event by how easy it is to recall instances of the event. Ease of recall is related to actual frequency, but is also related to several non-probabilistic factors such as the vividness or salience of the event, or how recently it occurred. These factors mean that the availability heuristic can give rise to biases such as the miscalibration of probabilities and the detection of illusory correlations.
3. *Representativeness* is the tendency to judge the probability of an event belonging to a class by the extent to which it is similar to or “representative of” that class. The classic demonstration of representativeness is the “Linda” effect in which a person (Linda) is usually adjudged to have a relatively high probability of being a librarian on the basis of a description that is “representative” of a stereotypical librarian [257]. Biases arising from the representativeness heuristic include ignoring or discounting base rate information and sample sizes, misconstruing chance events and sequences of random events, the conjunction fallacy, and failing to detect regression to the mean.

Although there is much empirical evidence in support of the heuristics and biases school (see the reviews listed above), there has also been strong criticism and contradictory evidence directed predominantly at the message that human judgement is inherently flawed. These rebuttals have received less attention than the original claims, but are well summarised in reviews in [134] and [157]. From the perspective of prescriptive decision aid, the most salient objections are that poor assessments are often found using experimental setups employing word problems that are strongly influenced by linguistic cues, and that these cues are often intentionally arranged to mislead or cause confusion for the subject. Simple changes to the structure of the problem can often greatly reduce the proportion of errors – for example, by using frequencies rather than probabilities [121], using negative framings [249], explicitly providing base rates [152], or making nested probabilistic structures explicit [235]. Johnson [134] cites evidence that “in favorable environments individuals can make excellent subjective probability judgments”. Favourable environments are those aided by expertise [135, 236], training and relevant feedback [180, 36], motivation [26, 13], a naturalistic rather than experimental setting [209, 134], and prediction tasks rather than memory retrieval tasks [286]. All of these are either common features of decision problems (e.g. naturalistic settings, prediction tasks) or would generally be considered good problem structuring practice (e.g. feedback, training, inclusion of relevant stakeholders representing experts and interest groups).



Shanteau [232] also lists characteristics of the kinds of tasks in which expert assessments tend to be good or bad. Many of the characteristics inducing good performance would be considered standard practice in prescriptive decision aid: decomposing the problem, having decision aids available, repetitive assessment so that consistency checks can be carried out, the incorporation of objective information where possible, and the provision of feedback. Evidence also suggests that assessments improve when decision problems are realistic and relevant to the decision maker [36]. Other aspects – the uniqueness of a task and a lack of inherent predictability – can lead to poor performance but simply reflect the complexity of typical multi-criteria applications. A fair summary seems to be that careful framing – how a question is asked as well as what is asked for – is a critical aspect of elicitation of any uncertain quantity, and that the mainstream heuristics and biases literature provides ample evidence of the types of errors that can be made if due care is not taken. However, the literature also indicates that prescriptive decision analysis may be relatively well-off; there is much evidence suggesting that good probability assessments can be obtained under the sorts of conditions which constitute current best practice.

An important question is to what extent these findings, which are almost all obtained from studies using assessments of probability, carry over to the other uncertainty formats discussed in this thesis. The different uncertainty formats described in Section 2.1.1 have attracted differing levels of attention from researchers, all dwarfed by the focus on probability, although there are some clear and valuable findings.

### **Explicit risk attributes**

Research on the assessment of explicit risk measures has focused on quantiles and ranges obtained from quantiles, and variances. Many of the findings relating to probability distributions apply to the assessment of quantiles as well. Experiments have found that people can often estimate quantiles reasonably well [102], although there is a clear tendency for people to overestimate their ability to predict an uncertain quantity so that their inter-quantile intervals tend to be too narrow [257, 174]. There is substantial uncertainty about which quantiles people are most accurate at assessing. Bisection methods can be used to obtain medians and upper and lower quartiles, with good results reported in [217] and [202]; other studies find that overconfidence is reduced if the 33% and 67% quantiles are used [103]. There is more agreement on the relative difficulty of accurately assessing the tails of the distribution [8], casting some doubt on the appropriateness of using extreme quantiles. As before, practice, training and feedback can all assist to improve the accuracy of assessments [102].

Variance, on the other hand, is a concept that people find difficult to interpret and to assess numerically [102]. Although empirical research is limited, results consistently in-

dicating substantial assessment errors. An early study demonstrated that assessments of the variance are often influenced by the mean [123], although even when the effect of the mean is removed assessments are poor [166, 27]. Variances, however, may be approximated from the quantiles using the results in [144, 82]. Means on the other hand can usually be assessed with a high degree of accuracy [28, 237], although if the distribution is highly skewed the estimate can be biased towards the median [216].

### **Fuzzy sets**

The assessment of fuzzy sets has received far less attention than it deserves given the subtlety of the ‘membership’ concept and the potential for multiple meanings. The lack of assessment procedures is a regular criticism of fuzzy decision aid (e.g. [100, 31]) which is discussed in more detail in Section 2.5, but here three points are worth making. Firstly, in the case of external uncertainty it is not uncommon for fuzzy sets to be treated in a very similar way to probability distributions: some authors advocate taking the possibility distribution i.e. the membership function, and standardising it so that the area under the function becomes one and it becomes a “probability” distribution (e.g. [127]). It might be expected that many of the heuristics and other findings that apply to probabilities also apply to possibilities. Certainly, since fuzzy membership denotes a degree of set membership, representativeness can be expected to play a role.

Secondly, many applications of fuzzy decision aid (e.g. [167, 96]) take the approach of defining triangular fuzzy numbers between two extreme points (e.g. the maximum, minimum, or some extreme quantiles) and one intermediate point (e.g. mode or median). Thus the same points noted for quantile assessment apply to these fuzzy numbers too. In particular, one might expect fuzzy numbers to have smaller supports than they ‘should’ i.e. to underestimate uncertainty. Other objections to the generic use of interval numbers (uniform distributions) and triangular distributions have been made in [207] and [102] – essentially claiming that these representations are in most cases overly simplistic.

Finally, fuzzy sets are often applied as a linguistic approach in which verbal assessments (e.g. “quite likely”, “fairly good”) are converted into fuzzy sets. Research into the verbal use of likelihood statements suggests that while people are more comfortable expressing their uncertainty verbally than numerically, the same phrase can be associated with very different probability ranges for different people, and these ranges can be strongly affected by contextual factors [175, 271, 199]. These findings all suggest that verbal assessments would need to be preceded by careful modelling of the meaning of the linguistic terms for them to be at all reliable.

## Scenarios

The assessment issues that arise when scenarios are used to represent uncertainty are somewhat more complicated because there are two types of assessments that must be performed. Firstly, there is the qualitative assessment of the scenarios themselves; then, the generally quantitative construction of alternatives within each scenario must also be assessed.

Most scenario planning texts deal explicitly with the question of scenario construction (e.g. [260, 230]), so that techniques for the development of scenarios are well-established. The traditional conditions for ‘good’ assessments are that scenarios should provide an internally consistent chain of cause-and-effect reasoning as well as offering novel insights into the decision problem (e.g. [260]). Clearly, these qualitative criteria make it far more difficult to judge consistency in the assessment of scenarios than for probabilities or other risk measures. Most studies into scenario assessment have dealt mainly with the issues of what constitutes good assessment [119] or why the current consistency criteria are sensible [288, 63]. The basic argument is provided in [288], which develops a theoretical model of the strategic process in which threatening and stressful circumstances – if unchecked – lead to ‘coping behaviour’ such as bolstering of the status quo and procrastination. This has a net result of entrenching a culture of conservatism and uncritical adherence. In this view, good scenario planning interventions overcome this tendency by providing challenging viewpoints and encouraging alternative views of what may happen. By implication, poorly assessed scenarios do not do these things, and rather become co-opted into the coping behaviour described above. Empirical support for this model has been reported in a review of an actual failed application of scenario planning [122].

There is thus some evidence to suggest the kinds of problems that can arise in actual scenario construction, although it seems that more research, and particularly more quantitative research, is needed in this area. Another matter is how the assessments within each scenario should be made, and what kinds of heuristics and biases may affect these assessments. To the best of my knowledge, there has been no research directly focusing on this area – something that can be attributed to the lack of formal decision analysis in strategic applications of scenario planning – but for the integration of scenario planning and decision analysis these seem critical. The few texts dealing with the assessment of information within scenarios typically use conventional elicitation approaches either within each scenario (i.e. under the conditions specified by that scenario) or simultaneously over all scenarios [109, 197]. This leaves open the key question of whether the nature of the underlying scenarios might bias evaluations in some way. For example, evidence suggests that people focus on optimistic scenarios and ignore pessimistic scenarios when predicting

their own performance, and also rate pessimistic scenarios as less likely to occur [203]. These differences might conceivably result in assessments in favourable scenarios being more reliable than those in unfavourable scenarios, or bias relative trade-offs between performance levels in different scenarios. Currently though, it is simply not known how and even which qualitative characteristics of scenarios, if any, influence the quantitative assessment of performance within scenarios.

## 2.2 Decision analysis with probability distributions

Modelling uncertainty about the consequences of actions using probabilities and probability theory represents what might be called the “classical” approach in decision analysis. This section presents three ways in which probability distributions have been employed in decision analysis: axiomatised multi-attribute utility theory, pairwise comparison methods, and methods using Monte Carlo simulation to draw from the distributions.

### 2.2.1 Multi-attribute utility theory

The multi-attribute version of utility theory is founded on the single-attribute case, so that it is useful to begin with the latter. The aim is to produce a utility function  $U$  satisfying the ‘expected utility hypothesis’: that an alternative  $a$  will be preferred to another alternative  $b$  if and only if the expected utility of  $a$  is greater than the expected utility of  $b$ . Von Neumann and Morgenstern [264] proved that such a function exists provided that a small number of superficially reasonable axioms are satisfied: completeness, transitivity, consistency, continuity, and independence. These axioms became the foundation of subjective expected utility theory. Within a relatively short space of time a number of systematic violations of these axioms, particularly independence, were reported (e.g. the famous paradoxes of Allais [7] and Ellsberg [94]). It is now a well-established fact that subjective expected utility does not adequately describe peoples’ choices, in the sense that a substantial proportion of people violate one or more of the axioms at least some of the time.

Violations of the axioms of expected utility elicit a number of responses which can be grouped into three broad categories. The first is that the axiomatic violations are ‘not that bad’, even in a descriptive sense: that the majority of decision makers are at least in tentative agreement with the axiom of independence [101] and that at least some apparent violations of expected utility theory can be classified as decision making with error rather than axiomatic violation [48]. The second is a pragmatic prescriptive response: that it is undoubtedly true that expected utility is descriptively invalid, and that violations are frequent and systematic, but that the theory retains a prescriptive value because it allows a simple and coherent framework for constructing preferences in a way that is transparent

and easy to explain and understand even to non-technical users (see chapter 4 of [31] for a defense of this argument). The final response is to seek to extend the expected utility model so that the resulting model is able to accommodate behaviour violating the axioms of expected utility. This has generated a large amount of interest and research for the case of a single attribute, although extensions for multi-criteria decision aid have only appeared more recently. These developments are the subject of Section 2.3.

In moving from one to several attributes, the aim of (now multi-attribute) utility theory remains to produce a function such that an alternative is preferred to another if and only if its expected utility is greater, but the presence of multiple attributes means that expectations must now be taken with respect to multivariate probability distributions. Practically, this requires (a) the construction of a marginal utility function  $u_j$  for each attribute  $c_j$ , and (b) some way of aggregating the marginal utility functions into a global utility function  $U$  satisfying the expected utility hypothesis. The existence of each individual marginal utility function depends on the satisfaction of the Von Neumann-Morgenstern axioms for each attribute. Suitably simple aggregation of the marginal utility functions requires that further assumptions hold – Keeney and Raiffa’s famous results showing that additive independence, mutual utility independence, and attribute-wise (but not full mutual) utility independence imply an additive, multiplicative, or multi-linear utility function respectively [146].

Some simplifications to the MAUT model have been proposed. Stewart [242] has shown that the effect of using an additive aggregation when preferences actually follow a multiplicative model is small over a range of problem settings. Von Winterfeldt and Edwards (in chapter 10 of [265]) indicate that there may be little difference in using marginal value functions instead of utility functions, which frees the decision maker from specifying preferences between lotteries. Some supporting evidence for this claim has recently been provided by Abdellaoui et al. [4] in the context of non-expected utility.

### 2.2.2 Pairwise comparisons of probability distributions

In some instances a pairwise comparison of probability distributions is sufficient to confirm that one alternative is preferred to another (in the sense of the expected utility hypothesis), provided that certain constraints on the underlying marginal utility function are satisfied. Three ‘stochastic dominance’ relations [114, 284] are defined as:

$$F_{a_j} >_1 F_{b_j} \iff H_1(x) = F_{a_j}(x) - F_{b_j}(x) \leq 0, \forall x \in [0, \infty) \quad (2.1)$$

$$F_{a_j} >_2 F_{b_j} \iff H_2(x) = \int_0^x H_1(y)dy \leq 0, \forall x \in [0, \infty) \quad (2.2)$$

$$F_{a_j} >_3 F_{b_j} \iff H_3(x) = \int_0^x H_2(y)dy \leq 0, \forall x \in [0, \infty) \quad (2.3)$$

where  $F_{aj}$  is the cumulative distribution function associated with  $Z_{aj}$  and the relation  $>_i$  refers to 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 stochastic dominance relations [25]. First-degree stochastic dominance implies that expected utility theory holds for all increasing utility functions; second-degree dominance implies it holds for all concave, increasing functions; and third-degree dominance implies it holds for all decreasingly risk averse, concave, increasing functions. Similar conditions have been provided for convex utility functions [110, 299]. When moving into a multicriteria framework, Huang et al. [126] have shown that a necessary condition for multiattribute stochastic dominance is stochastic dominance on each individual criterion, although the conflicting nature of multicriteria problems means that this is perhaps unlikely to occur often.

Pairwise comparisons of probability distributions have also been incorporated into a number of stochastic outranking methods. Jacquet-Lagrèze [130] removes the intersection of the two probability mass functions as evidence in support of indifference, and then uses the cumulative distributions to allocate the remaining probability mass as evidence either that alternative  $a$  is preferred to  $b$ , or *vice versa*. Aggregation proceeds as for ELECTRE I. A second set of models [78, 187, 188] 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  $c_j$  i.e.  $\Pr[Z_{aj} \geq Z_{bj}]$ . The models differ with respect to the subsequent exploitation of the probabilities. Dendrou et al. [78] simply aggregate the  $P_{ab}^j$  using a weighted sum over criteria to arrive at a global index for each pairwise comparison  $P_{ab}$ . Martel et al. [187, 188] incorporate more sophisticated outranking concepts such as indifference and preference thresholds, and subsequent aggregation and exploitation proceeds in a similar fashion to ELECTRE III. In all of these models the distributional aspect of the problem is fully absorbed into the problem at an early stage of the process through the definition of the  $P_{ab}^j$ . In contrast, d'Avignon and Vincke [73] use the uncertain attribute evaluations to form a stochastic (or 'distributive') outranking degree indicating the probability of attaining various degrees of outranking, rather than summarising the stochastic evaluations directly as  $P_{ab}^j$ . A fourth group of models [301, 300, 188, 14, 205] uses stochastic dominance relations. Zaras and Martel [301, 300] use a simple weighted aggregation of indicator variables  $c_j(a, b)$  which equal 1 if  $a$  stochastically dominates  $b$  on criterion  $c_j$  and are otherwise zero. This results in a concordance index as for ELECTRE I. Martel et al. [188] and Azondékon and Martel. [14] use a more nuanced approach by defining the local preference index  $c_j(a, b)$  as a product of three functions each scaled between 0 and 1 that cause  $c_j(a, b)$  to decrease as dominance conditions weaken from first- to third-order stochastic dominance. A similar threshold-based method is provided in [205].

### 2.2.3 Models simulating from probability distributions

If evaluations can be specified using probability distributions, one option is to use Monte Carlo simulation to generate values from the distributions and to use these simulated values as inputs to a decision model. This approach has been particularly popular in the analytic hierarchy process (AHP). Early research into the modelling of probabilities in the AHP was largely concerned with deriving relationships between the distributional form of the uncertain pairwise judgements and the distributions of the marginal evaluations contained in the ‘priority vector’ i.e. the eigenvector corresponding to the largest eigenvalue [262, 22, 23]. Subsequent probabilistic AHP models [170, 171, 24, 18] have focused on using Monte Carlo simulation to randomly generate pairwise evaluations from the distributions specified by decision makers. These approaches all follow the same basic approach. The decision maker expresses pairwise comparisons in the usual way i.e. using the same 1–9 scale as for deterministic AHP, except that these comparisons are allowed to be random variables with associated probability distributions. No restriction is placed on the types of distributions that are appropriate. Next, sets of random pairwise judgements are generated using Monte Carlo simulation. For each set of randomly generated evaluation matrices, the priority vector is computed. Repeating this process many times gives a distribution of priorities for each alternative. These simulated distributions can be used to rank the alternatives; in most cases using the mean of the distribution.

Most authors make small embellishments around this general process. Levary and Wan [170, 171] incorporate scenarios into their model. Decision makers thus assess different (possibly stochastic) judgemental matrices for each scenario. Their simulation approach first generates a random number to specify which scenario is being used, and then generates further random numbers specifying the pairwise judgements within each scenario. Basak [24] uses a Bayesian approach to integrate expert judgements with the decision maker’s prior probabilistic assessments. Pairwise judgements are simulated by drawing from the posterior distributions. Bañuelas and Antony [18] use the basic simulation approach but add a more substantial *post hoc* sensitivity analysis phase to investigate the most important sources and effects of the uncertainty.

The other area of MCDA in which Monte Carlo simulation from probability distributions has been popular is in the family of stochastic multicriteria acceptability analysis (SMAA) methods. The SMAA methods are essentially inverse-preference methods that are useful in situations where it is difficult or impossible for a decision maker to explicitly give preference information, for example in highly antagonistic political situations. In these cases, information is provided about the types of preferences that would lead to the selection of particular alternatives, given the information provided. This is accomplished by simulating

from distributions defined over the values of possible attribute evaluations and preference parameters. The models thus incorporate both internal uncertainty (about preferences) and external uncertainty (about attribute evaluations). The extent to which decision makers may express their preferences can be located anywhere along a continuum from no assessment at all [125] to complete assessment of preferences as for regular prescriptive decision aid [141, 251]. The two key outputs of a SMAA analysis are the acceptability index, which is the relative proportion of all simulations in which each alternative attains a particular rank (most importantly, the top rank), and the central preference vector, which summarises the types of preferences that lead to a particular alternative being the preferred one. Most often this is computed using the center of gravity of the hypervolume containing all the preferences supporting the selection of an alternative. Further details are provided as part of the applications in Chapter 5.

The original SMAA method in [159] analysed the combinations of attribute weights that result in each of a set of prospective alternatives being selected when using an additive utility function. Subsequently, the SMAA methodology has been extended to several other prescriptive models of choice: to outranking in [250], to goal programming and reference point methods in [161] and [87], to data envelopment analysis in [163], to prospect theory in [164]; and to ordinal and mixed ordinal-cardinal evaluations in [165, 160]. Although it appears not to have been directly applied to the AHP, it is clear that such an application would result in a model very similar to those discussed above. Practical examples of the application of SMAA in a prescriptive context are given in [141] and [251], while two recent studies have also applied SMAA in a descriptive setting [88, 89]. Further technical details about implementing SMAA can be found in [252].

## **2.3 Decision analysis with decision weights**

Whether or not simplifications to multi-attribute utility theory lead to worse prescriptive decisions, the results obtained by Allais [7], Ellsberg [94] and later many others indicate that expected utility theory does not accurately represent actual i.e. descriptive, choices. This has led others to develop models of choice that are able to capture the elements of choice under uncertainty that expected utility cannot. While these are predominantly descriptive theories, their influence has extended into normative and prescriptive modelling in the search for models that are “more responsive to the complexities and limitations of the human mind” [256].

### **2.3.1 Empirical evidence**

Although early empirical evidence such as the ‘paradoxes’ of Allais and Ellsberg was aimed at highlighting the descriptive failings of expected utility, and thus on how decisions were



*not* made, there is now a fairly good idea of what properties a decision model should possess if it is to be descriptively valid. Starmer [239] presents a comprehensive review of results obtained from a large number of experimental studies using probability triangles,<sup>1</sup> from which he draws three “stylised facts”. The first is that although linear parallel indifference curves (as predicted by expected utility theory) are inappropriate, so too are generalised fanning-out (relative to an origin outside the triangle) indifference curves. Fanning out implies that decision makers become more risk averse as their prospects improve but empirical evidence suggests that they in fact become less risk averse (e.g. [67, 45]). Descriptive decision models require a mixture of fanning-out and fanning-in.

Secondly, indifference curves should not be required to be linear. Linearity imposes the requirement that a mixture of two gambles cannot be preferred to either of those gambles, and that a decision maker should be indifferent between any mixture of two gambles that are equally valued. These requirements constitute a weak form of the independence axiom called ‘betweenness’. Empirical evidence suggests that decision makers are often unwilling to randomise equally-valued gambles [47], implying quasi-convex rather than linear indifference curves.

A third empirical observation is that violations of expected utility theory tend to be more pronounced when certain or near-certain prospects are involved i.e. on or near the boundaries of the probability triangle. This does not imply that expected utility theory is conformed to in the interior of the triangle (see [289] for violations of that sort), just that behaviour is less conforming on the boundaries (e.g. [71, 46]). The fact that decision makers weigh extreme probabilities not just by their probability of occurrence – and are more sensitive to changes in the probability of an outcome with an extreme probability – suggests the use of a non-linear function to weigh the probabilities according to their importance to the decision maker.

Some further empirical observations have been added by the work of Kahneman and Tversky (e.g. [139, 258]), most of which have been incorporated into their prospect theories, which are discussed in the next subsection. These can be summarised as (a) the presence of a reference point relative to which outcomes are evaluated as gains or losses, (b) loss aversion, the tendency for losses to “loom larger than corresponding gains” [258], (c) diminishing marginal sensitivity to both gains and loss, and (d) non-linear weighting of probabilities, in particular one which underweights large probabilities and overweights

---

<sup>1</sup>Probability triangles are graphical representations of gambles (or lotteries) defined over three outcomes  $x_1, x_2, x_3$  with  $x_1 < x_2 < x_3$ , so that any gamble can be represented on the triangle using the probabilities of each outcome occurring,  $(p_1, 1 - p_1 - p_3, p_3)$ . Lines of indifference can be drawn on the triangle between gambles a decision maker is indifferent between, and it is the pattern of these indifference curves that are a major focus of empirical research – because different decision model make different predictions about the properties the indifference curves possess.

small probabilities. The first three of these observations suggest utility functions that are convex for losses and concave for gains, and are steeper in the domain of losses; the last observation suggests an inverted s-shape form for the probability weighting function.

Empirical observations are useful because they allow entire classes of models to be evaluated as appropriate or otherwise. As discussed in the next section, the observations above leave relatively few descriptively-sound candidates. Normatively and prescriptively, of course, one must evaluate whether the observations indicate traits that one would wish to include for good decision making. This is done as part of Section 2.3.3; before that I review the non-expected utility theory models.

### 2.3.2 Descriptive theories of choice under uncertainty

The earliest violations of expected utility theory (e.g. Allais' paradox [7]) can be explained with models possessing indifference curves that fan outwards. Models of this type include those using (a) weighted utility theory [64], which replaces the independence axiom with a weaker independence requirement which constrains indifference curves to be linear but not necessarily parallel, and (b) the theory of disappointment [30], which has not been axiomatised but values prospects using expected utility as well as an additional expected 'disappointment' or 'elation' caused by an outcome being respectively worse or better than prior expectations.

Empirical evidence suggests that indifference curves fan both inwards and outwards (in the same probability triangle). This type of behaviour can be accommodated by assuming betweenness rather than independence, an approach followed by implicit expected utility theory [77] and implicit weighted utility theory [65]. Betweenness still implies linear indifference curves though (the converse is also true), which is contrary to empirical evidence suggesting a regular aversion to randomisation of equally-valued lotteries. Thus betweenness must be relaxed as well. This has been done in models using quadratic utility theory [66], which replaces betweenness with the weaker requirement of mixture symmetry<sup>2</sup>, and the even less restrictive gamble dependent utility theory [29], which assumes only ordering, continuity and monotonicity and thus makes nearly no predictions of behaviour at all.

Other non-expected utility theories assign subjective 'decision weights' to the probabilities of different consequences, rather than assigning weights to the consequences themselves (as the models discussed above do). These models usually do not satisfy betweenness because the sum of the decision weights of complementary events may be sub- or super-additive depending on the form of the weighting function. In a descriptive setting, this is a distinct

---

<sup>2</sup>Mixture symmetry holds if, for two equally-valued lotteries  $\mathbf{x}$  and  $\mathbf{y}$ , the decision maker is indifferent between the mixture lottery with probability  $p$  of obtaining  $\mathbf{x}$  and  $1 - p$  of obtaining  $\mathbf{y}$  and one with probability  $1 - p$  of obtaining  $\mathbf{x}$  and  $p$  of obtaining  $\mathbf{y}$

advantage since it allows for non-linear indifference curves, as empirical evidence suggests is necessary. The so-called simple decision weighted utility models – in which decision weights take the place of probabilities in the traditional expected utility formulation – have the undesirable consequence that the resulting preference function is non-monotonic and dominated options may be selected. This problem is avoided by rank-dependent expected utility [218, 240], which assigns decision weights based on both the probability of an outcome as well as its rank relative to others in magnitude, and cumulative prospect theory (without an editing phase) [258], which uses a utility function employing reference points, loss aversion, and diminishing marginal sensitivity, and an inverted s-shaped function to transform cumulative probabilities into decision weights.

Regret theory [183] evaluates prospects relative to the possible consequences of other choices rather than to a fixed reference point. The utility of an outcome is measured by a function that increases in the value of the outcome and decreases in the value of outcomes on other options. When prospects are statistically independent from one another, regret theory is a general case of the weighted utility theory discussed earlier, in which the axiom of transitivity has been relaxed. It therefore predicts specific types of cyclical or non-transitive choices. Some of these predictions are supported by empirical evidence [181, 182], but other studies have found empirical non-transitive choices that cannot be accommodated by regret theory [241].

### **2.3.3 Developments in prescriptive multi-criteria decision aid**

Traditionally debate about the prescriptive usefulness of non-expected utility theories has tended towards philosophical argument about the normative status of the various violations of expected utility (see [9] for a review as well as a defence of the normative validity of many of the violations). As pointed out in [268] however, a motivation for non-expected utility theories can be found without entering this debate. A key part of the practical implementation of utility theory is the elicitation of utilities from the decision maker. A common assumption of many elicitation techniques (e.g. the certainty-equivalent and probability-equivalent methods) is that utilities can be inferred from information provided by the decision maker using the tools of expected utility theory. Bleichrodt et al. [34] call this the “classical elicitation assumption”, and make the important points that (a) this is a descriptive assumption, dealing with observed behaviour, and is thus independent of the ‘normative’ debate; (b) imposing the assumption on a decision maker who does not follow expected utility theory can lead to biased assessments of utility. Some inconsistencies may be resolved through discussions between the decision maker and analyst – this is the approach advocated as ideal in [34] – but there will be others that are resistant to such discussion. For such cases, utilities assessed using expected utility theory will provide

faulty inputs to later modelling phases – whatever form this modelling may take.

The practical problems involving the prescriptive use of expected utility theory have led to some exciting developments in the integration of prescriptive decision aid and non-expected utility. These can be divided into three broad areas. The first is the development of alternative assessment techniques that can be used to construct utility functions without using expected utility foundations i.e. that can construct non-expected utility functions. The second is the empirical question of how much the classical elicitation assumption biases the assessment of utilities, and the consequential effect on decisions. The third is the development of procedures for the decomposition of multi-attribute non-expected utility functions, similar to those formulated by Keeney and Raiffa [146] for multi-attribute (expected) utility theory.

Significant advances have been made in all three of these areas. Wakker and Deneffe [268] propose a utility assessment method that does not depend on the actual probability values, and is thus robust to the kinds of probability transformations that decision makers often use. Their *gamble-tradeoff method* uses two reference outcomes  $R$  and  $r$  (with  $R > r$ ) and elicits the value  $X$  that makes a decision maker indifferent between the gambles<sup>3</sup>  $(X, p; r)$  and  $(x, p; R)$  for some  $x$ , and the value  $Y$  for which  $(Y, p; r) \sim (y, p; R)$  for some  $y$ . It can be shown that this pair of indifferences can be rewritten in the form  $u(X) - u(x) = u(Y) - u(y)$  and thus that a standard sequence of indifference assessments will yield a utility function without needing to know the value of the probability  $p$ . Wakker and Deneffe suggest that events without known probability can be used in place of  $p$  e.g. “surgery will succeed” [268]. All that matters is that the decision maker uses  $p$  consistently throughout. Extensions by Abdellaoui [3] and Bleichrodt and Pinto [33] allow for the assessment of both the non-expected utility function and the probability weighting function. Both methods employ the utilities obtained with the gamble-tradeoff method to elicit decision weights using a set of further indifference questions. Bleichrodt et al. [34] develop standard correction procedures for the certainty and probability equivalence methods for cases in which time or other resource constraints prevent interactive discussions between decision maker and analyst, using median probability weighting functions obtained from Kahneman and Tversky’s work on cumulative prospect theory [258]. More recently, further extensions have been proposed in [5] that allow the full assessment<sup>4</sup> of the prospect theory utility function i.e. one that is defined over the whole domain of losses and gains.

---

<sup>3</sup>Notationally I write a gamble returning an outcome  $X$  with probability  $p$  and an outcome  $r$  with probability  $1 - p$  as  $(X, p; r)$ .

<sup>4</sup>Previous methods, like [268], assessed the prospect theory utility function separately for gains and losses, because under prospect theory the probability transformation function may be different in the loss and gain domains. Abdellaoui et al. [5] use a series of three indifference questions to link the utility on the loss domain with the utility on the gain domain, and thus assess the entire utility function.

All of the above-mentioned papers contribute empirical results in addition to their practical developments. Utility functions obtained using assessment procedures that are sensitive to probability weighting exhibit more risk aversion for gains [268, 5] and more risk proneness for losses [5] than utility functions obtained with the gamble-tradeoff method. Wakker and Deneffe [268] interpret this as a result of certainty effects. The corrective procedures in [34] make fairly large adjustments to both certainty and probability equivalence methods, particularly at high probabilities and particularly for the probability equivalence method. The greater sensitivity of the probability equivalence method – also reported in [268] – is ascribed to response-mode effects in which the use of a probability scale in the former method draws attention to that aspect of the problem [268]. Most evidence supports the form of the probability weighting function proposed by prospect theory: an inverted s-shape for the probability weighting function [3, 33], and different weighting functions for losses and gains [3]. Linearity at intermediate probabilities (implying closer correspondence with expected utility theory) is supported in one study [33] but refuted in another [3]. The only model able to elicit full prospect theory utility functions [5] finds utility functions that agree with the theory (convex for losses and concave for gains) for just over half of the subjects taking part. Perhaps the most intriguing results from a practical perspective are found by Abdellaoui et al. in [4]. They compare a number of methods for assessing utility functions and find that (a) there are no inconsistencies between assessment methods once probabilities have been transformed according to prospect theory, and (b) there is no difference between the value functions elicited using riskless strength-of-preference information and the utility functions elicited using more complex choices between lotteries. The latter finding in particular seems important for practical decision aid.

Multi-attribute applications require some way of decomposing the multi-attribute preference function into its simpler marginal constituents. In expected utility theory, Keeney and Raiffa [146] were able to use various independence conditions to provide a number of different decompositions – in particular the multi-linear, multiplicative, and additive representations. Miyamoto and Wakker [194] show that many of the same decomposition results obtained under expected utility theory are also valid under non-expected utility. The decomposition procedures use the same definition of utility independence used by Keeney and Raiffa, except that the condition is defined over the set of all rank-ordered gambles<sup>5</sup> includes rather than *all* gambles. Utility independence of each attribute and full mutual utility independence are necessary and sufficient to infer multilinear and multiplicative aggregations respectively, as in expected utility theory, but the conditions for an

---

<sup>5</sup>Gambles whose outcomes have been ranked in descending order of preference i.e. a two-outcome gamble  $(x, p; y)$  is rank-ordered if  $x \succeq y$ . Note that the set of all rank-ordered gambles contains all deterministic outcomes.

additive representation are somewhat different. This is because marginality – the condition that preferences between gambles depends only on the marginal distributions – imposes linearity in the probabilities and has thus been shown to imply an expected utility model [92]. Independence results have been used to provide multi-attribute representations for (among others) rank-dependent expected utility and prospect theory in [194], and cumulative prospect theory in [298] and [32]. These developments in non-expected utility suggest that multi-attribute modelling using non-expected utility foundations would not look a great deal different from that done using expected utility.

## 2.4 Decision analysis with explicit risk attributes

Given the aim of taking external uncertainty about outcomes on an attribute into account, one possible approach is to use some measure of this uncertainty as an attribute in its own right. This approach provides a single (or small number of) risk measures indicating how variable or ‘risky’ performance is. The fundamental notion is that uncertain evaluations can be expressed in terms of ‘value’ and ‘risk’ components. Multi-attribute risk-value models take the form

$$U_i^{(\text{risk})} = \sum_{j=1}^J \left[ w_j u_j(V_{ij}) - \sum_{k=1}^{\mathcal{K}} w_{ij}^{R^{(k)}} R_{ij}^{(k)} \right] \quad (2.4)$$

where  $V_{ij}$  and  $R_{ij}^{(k)}$  are measures of the ‘value’ and ‘risk’ of  $Z_{ij}$  respectively,  $w_j$  is a ‘swing weight’ reflecting the relative importance of a one-unit change in attribute  $c_j$ , and  $w_{ij}^{R^{(k)}}$  is a weight for  $R_{ij}^{(k)}$ , termed a ‘risk weight’ and also interpreted as a swing weight. Note that in this general formulation the risk weights may depend on alternatives as well as attributes. The number of risk measures used is denoted by  $\mathcal{K}$  but usually there will only be a single measure and the superscript  $k$  will be dropped. The measurement of value is relatively uncontroversial and the use of expected values for this purpose i.e.  $V_{ij} = E[Z_{ij}]$ , is widely accepted [228]. There is far less agreement on an appropriate measurement of uncertainty, and there remain several conflicting notions about how uncertainty should be defined and modelled [98].

### 2.4.1 The measurement of risk

One way of approaching the problem of risk measurement is to attempt to describe what people mean when they say that an event is ‘risky’. Weber and Bottom [282] provide a review of empirical research on how the attributes of gambles influence risk and conclude that: 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; and risk increases if a gamble is repeated many times. These are descriptive points, and need to

be interpreted carefully for the purpose of any normative or prescriptive modelling. Risk attitudes and in particular judgements involving probabilities are, as discussed in Section 2.1.2, notoriously prone to biases that can lead to what in many circumstances would be considered poor decisions.

Sarin and Weber [228] and Jia and Dyer [133] consider links between measures of risk and preference models in the context of (single-attribute) lotteries. Variance can be used as a risk measure if the decision maker has a quadratic utility function or where the utility function is increasing and concave and the random variables representing lotteries are normally or log-normally distributed. A linear combination of variance and skewness can be used as a risk measure if the utility function is a third-order polynomial, while another risk measure,  $bE[e^{c(X-\bar{X})}]$ , is appropriate for a utility function of the form  $U(x) = ax - be^{-cx}$ . For any multi-criteria modelling, two issues are of special importance. As Stewart points out in [246], it seems doubtful that decision makers would be able to give meaningful answers to trade-off questions based on anything more complex than the variance or standard deviation. In the multi-criteria problem it would be very difficult to use different risk measures for different attributes (where the underlying marginal utility functions are of different functional forms). It is doubtful that this additional complexity would add any insight for the decision maker, and the scope for confusion seems considerable.

An alternate approach is to consider risk as the probability of ‘something bad happening’ – more formally, of some target not being met. This has intuitively appealing associations with everyday usage of the term ‘risk’, and evaluating alternatives by the chance of a resulting catastrophe has a well-established history in risk analysis (e.g. [97]). In a multi-criteria context, Stewart [245] found that decision makers were able to think clearly about risk in terms of the probabilities of fishing stocks lasting certain periods of time before collapsing. Other researchers have also proposed the probability of loss as a measure of risk [98], as well as a non-linear combination of the probability of losses and gains [185]. However the use of a single fixed target means that there is no guaranteed existence of an equivalent utility function formulation for these probability-based risk measures [49]. In order for such an equivalence to exist, the target must be probabilistic. Castagnoli and Li Calzi [49] show that an alternative formulation of the expected utility model is to assume a decision maker who has only two different utility levels depending on whether an uncertain target is met or not. The ‘target oriented’ decision maker assesses probabilities  $p(x)$  that the target is achieved given an attribute performance of  $x$ , rather than a utility function  $u(x)$ . It is argued in [38] that in some circumstances this may be a “more intuitively appealing task”.

### 2.4.2 Multi-attribute applications

Stewart [245] applies both variances and probabilities of poor performance to a fish-stock management problem. That study (a) showed that the probability of poor performance – collapse of a fish population over various time horizons – can provide a representation of uncertainty that decision makers find useful to work with, (b) argued that the variance may sometimes be a misleading risk measure (when the utility function took the form of a power function and lotteries were non-normal) but in other circumstances may be perfectly adequate (when an exponential utility function was used, even where lotteries were non-normal), and (c) suggested that “the use of two well-chosen cumulative probabilities may be useful” (again with mixed results depending on the form of the underlying utility function).

Bordley and Kirkwood [38] extend the single-attribute results in [49] to show that for each multi-linear (or multiplicative or additive) utility function, there is an equivalent multi-linear (or multiplicative or additive) target-oriented formulation. Some basic assessment procedures are provided, and it appears that weight elicitation could technically proceed using the usual SMART procedure (e.g. [265]), although it remains an open question whether decision makers would find such information easy or useful to consider.

Further detail on the use of variances is provided by Kirkwood [150]. He shows that using (2.4) with  $V_{ij} = E[Z_{ij}]$ ,  $R_{ij} = \sigma_{ij}^2$ , and  $w_{ij}^R = (1/2)w_j u_j''(E[Z_{ij}])$  can lead to close approximations of expected utility under the important conditions that the  $Z_{ij}$  be normally distributed or numerous enough for the central limit theorem to apply, and the underlying utility functions  $u_j$  “do not deviate too much from linear”. Results from another simulation study ([91]; also presented as part of Chapter 3 of this thesis) found poor approximation accuracy under strongly non-linear preferences, and that in such situations a model using only expected values i.e. using  $w_{ij}^R = 0$ , performed better. Further examples of the use of explicit risk attributes in multi-attribute analysis can be found in the goal programming methods, both of the variance-based [83, 15, 16] and probability-based [206, 149, 75] variety. Both the variance-based and probability-based goal programming models can be shown to be special cases of the target-oriented preference model [38]. Hallerbach and Spronk [115] use variances to measure risk in their ‘multiple factor approach’ to portfolio selection.

### 2.4.3 Models using quantiles

In practical decision analysis it is common to represent probability distributions using a small number of (usually three to five) quantiles [102], this being the basis for well-known elicitation methods like the bisection or interval methods [e.g. 238]. Keefer and Bodily



[144] have also shown that the single-attribute expected utility of an alternative can be closely approximated by assessing utilities at each of the 5%, 50%, and 95% quantiles of performance and forming a weighted sum in which the median receives a weight of 0.63 and the 5% and 95% quantiles each receive a weight of 0.185. Triples consisting of the minimum, median/mode, and maximum are also popular in fuzzy decision analysis [e.g. 167, 96]. Obtained quantiles may be used in several ways – full distributions may be fitted to the quantiles, they can be used to provide moment estimates using approximations such as those in [144], or they can be considered as criteria in their own right. Figure 2.1 shows a value tree formulation for a hypothetical decision problem consisting of three uncertain criteria which are represented by their 5%, 50%, and 95% quantiles.

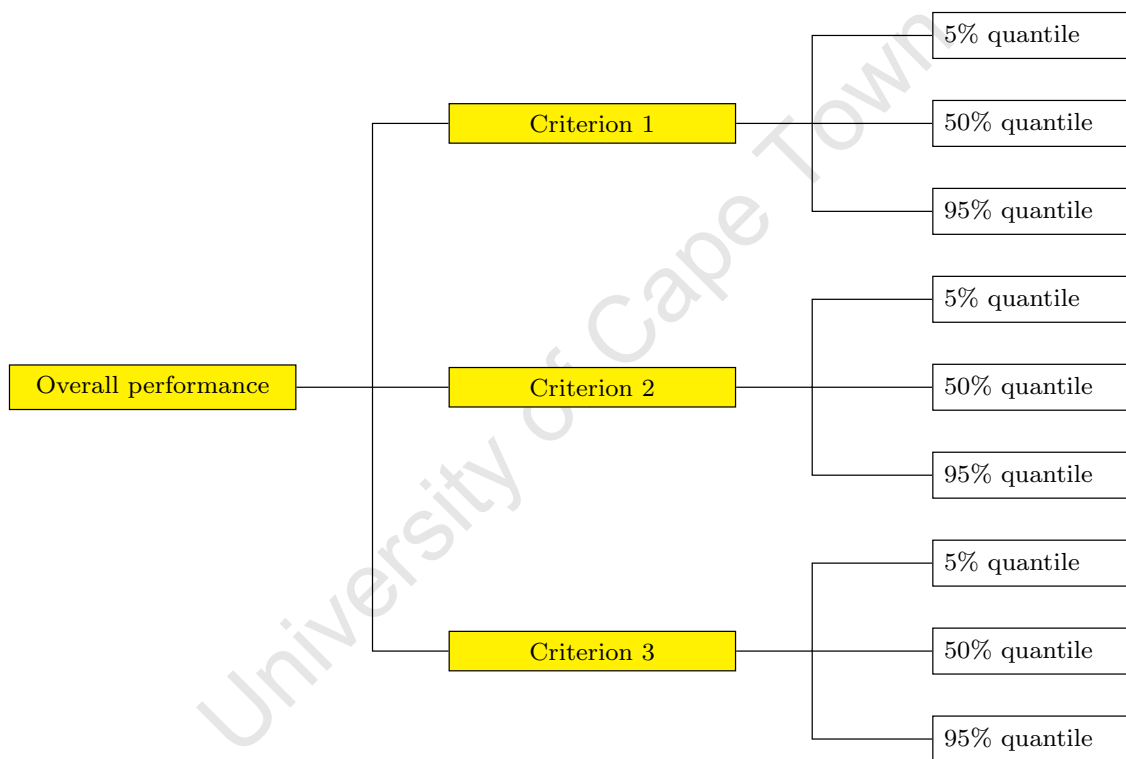


Figure 2.1: A value tree illustrating a hypothetical MCDA process where uncertain evaluations are represented using quantiles.

Figure 2.1 represents what might be called a ‘conventional’ view of quantile-based MCDA, in which uncertainty is resolved at the lower-levels of the value tree. An alternate approach is to place the quantiles at the second level of the tree, as shown in Figure 2.2. This has the effect of creating a ‘super-MCDA’ problem consisting of  $N_q$  ‘deterministic’ and generally closely related problems, where  $N_q$  is the number of quantiles used. Aggregation over quantiles is therefore carried out at the end of the decision process rather than at the beginning. This allows decision makers to compare alternatives separately at their (effective) worst, intermediate, and best levels. Although this is a somewhat unusual

point-of-view for uncertainty modelling using quantiles, it is similar to the approach taken by scenario-based MCDA, which is described in Section 2.6 (although a single quantile cannot generally be described by an internally consistent ‘storyline’, as a scenario can). It is conjectured that decision makers may find the additional information provided within each quantile useful (Chapter 5 describes an application where this is the case).

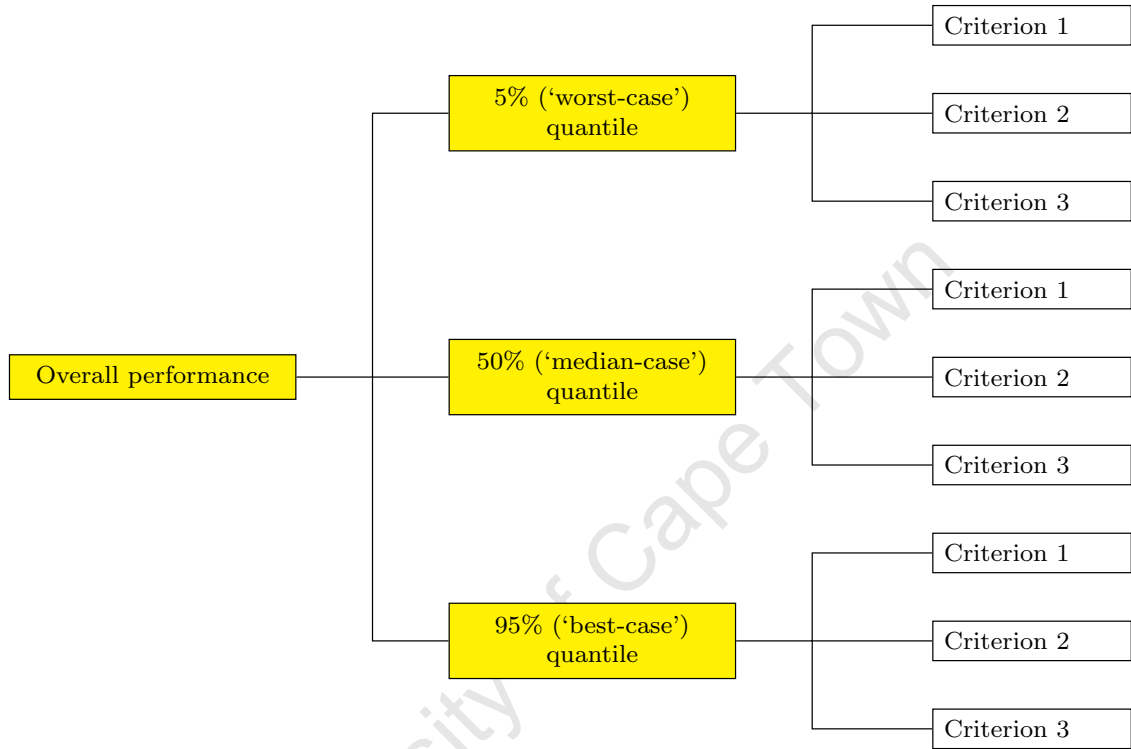


Figure 2.2: An alternate value tree formulation for quantile-based MCDA. In this tree, quantiles are represented in the second level of the objectives hierarchy rather than as lower-level criteria.

The value tree formulations in Figure 2.1 and 2.2 are mathematically equivalent, so that in either case the general quantile model evaluates alternative  $a_i$  by

$$U_i^{(\text{quan})} = \sum_{k=1}^{N_q} \sum_{j=1}^J w_{jk} u_{jk}(z_{ij}^{(q_k)}) \quad (2.5)$$

where  $q_k$  refers to a specific quantile,  $z_{ij}^{(q_k)}$  is the  $q_k$ -th quantile of  $Z_{ij}$ , and  $w_{jk}$  and  $u_{jk}$  are respectively the weight and marginal utility function associated with quantile  $q_k$  and attribute  $c_j$ . Note that in this general formulation the attribute importance weight and marginal utility function associated with a particular criterion  $c_j$  is allowed to differ over quantiles  $q_k$ . Where necessary (which would perhaps not be often) joint weights can be evaluated in the usual way for higher-level criteria. First, relative criterion weights  $w_{j|k}$  can be assessed at a particular quantile (following the notation in [31], the relative weight  $w_{j|k}$

denotes the weight of criterion  $c_j$  at quantile  $q_k$ , with  $\sum_j w_{j|k} = 1, \forall k$ ). Then, quantile weights can be assessed i.e. weights  $w_{q_k}$  associated with each quantile  $q_k$ , and joint weights are given by the product of the two i.e.  $w_{jk} = w_{q_k} w_{j|k}$ . The practical interpretation and assessment of the quantile weights  $w_{q_k}$  has not been fully resolved and requires some care. On one hand, the Keefer-Bodily values indicate values for  $w_{q_k}$  which, on average, give the best approximation of expected utility. On the other hand, since the  $\sum_{j=1}^J w_j u_j(z_{ij}^{(q_k)})$  values constitute an interval preference scale, the  $w_{q_k}$  can also be interpreted as relative (“swing”) weights on performance in different quantiles. These could be assessed based on the usual trade-off questions involving pairs of quantiles on a common attribute, and there is no theoretical reason why in general they should be equal to the Keefer-Bodily weights. At present there is little guidance on which approach is to be preferred.

Alternatively the  $w_{jk}$  can be assessed directly using comparisons between performance levels on different combinations of attributes and quantiles, although this may be impractical for large numbers of attributes or quantiles. In the perhaps more conventional case where weights and quantiles do not differ over quantiles alternative  $a_i$  can be evaluated by

$$\sum_{k=1}^{N_q} \left[ w_{q_k} \sum_{j=1}^J w_j u_j(z_{ij}^{(q_k)}) \right] \quad (2.6)$$

Decision models using quantiles can be expressed as explicit risk models with the value component set to the median rather than the mean i.e.  $V_{ij} = z_{ij}^{(0.5)}$ , and extreme quantiles used to capture the ‘risk’ components e.g.  $R_{ij}^{(1)} = z_{ij}^{(0.05)}$  and  $R_{ij}^{(2)} = z_{ij}^{(0.95)}$ , with  $w_{ij}^{R^{(k)}} = -w_j w_{q_k}$  and  $q_1$  and  $q_2$  referring to the 5% and 95% quantile respectively. For this reason they have been included here under the heading of the explicit risk models. Two example applications can be found in [245].

## 2.5 Decision analysis with fuzzy sets

Fuzzy decision models seek to model the elements of the decision-making process that are subject to uncertainty using fuzzy sets. From a philosophical point-of-view, uncertainty in any of these elements is said to arise because of (a) unquantifiable information, (b) incomplete information, (c) nonobtainable information, and (d) partial ignorance [59]. Practically, there is usually no difference in how the types of uncertainty are incorporated into a fuzzy decision model. Fuzzy multi-attribute decision making is usually divided into two stages [85, 303, 59]: a first stage consisting of the assessment and aggregation of attribute evaluations; and a second stage producing a rank order of alternatives from the aggregated performance ratings of the first stage. Each of these stages is discussed in turn in the remainder of this section.

### 2.5.1 Aggregation models for fuzzy evaluation

Many if not all decision models can be fuzzified because of the existence of fuzzy versions of nearly all operations that are likely to be employed by a decision model (e.g. addition, multiplication, finding a minimum or maximum). Three decision models that have received the most attention from fuzzy researchers are focused on: weighted additive models, models based on ideas related to the analytic hierarchy process, and aspiration-based models using comparisons to constructed ideal and anti-ideal solutions.

#### Models using weighted additive sums

Fuzzy weighted additive aggregation models have been used in [54, 51, 70] and [120]. The first three of these papers differ mainly with respect to the way in which the fuzzy numbers formed using the weighted additive aggregation are exploited to arrive at a ranking. The final paper also employs a simple weighted additive model, but this time with unknown weights. A linear programming formulation is used to solve for the weights under the somewhat strange assumption that one wishes to maximise the sum of absolute weighted evaluation differences over all alternatives and attributes – on the basis that this maximises some kind of collective ‘importance’. The case of solving for unknown attribute weights is fairly popular (see also [172] and [278], which also use a weighted additive model) but it is outside the scope of the current thesis, which considers the more traditional context in which weights would be assessed as part of the decision aiding process.

#### Models using the analytic hierarchy process

The decision aiding approach that seems to have received the most attention from fuzzy practitioners is the AHP. On the face of it, the AHP seems an obvious candidate for fuzzification because of the qualitative nature of the pairwise comparisons made. Although these judgements are traditionally quantified on a 1-9 scale, the points along the scale are given linguistic meanings and comparisons will often be made predominantly with these linguistic labels in mind. It is therefore easy to imagine that a decision maker might be uncertain about some of these judgements. Since the 1996 reviews, I have found more than 30 papers involving the application of fuzzy set theory to the AHP, and there are doubtless many more. The methodological papers differ predominantly in how the vector of marginal preference information is computed from the matrix of judgemental ratios assessed by the decision maker.

It is well-known in the AHP literature that because the judgements expressed by the decision maker may not be consistent (in the sense that  $r_{ik} = r_{ij}r_{jk}$ , where  $r_{ik}$  denotes the (ratio-scaled) strength of preference of  $a_i$  over  $a_k$ ), there are a variety of ways to estimate the marginal information. The original approach (e.g. [225]) uses the principal

eigenvector of the matrix of assessed judgements, but two further estimation approaches are also popular, using least squares (e.g. [132]) or logarithmic least squares optimisation (e.g. [156]). Exactly which method is best has been a point of some debate over the years (e.g. [226]). All three estimation approaches have been modified to make use of fuzzy input data: eigenvalue methods in [254, 72, 274], least squares in [291], and log least squares in [261, 35, 292, 276]. Several other methods have also been proposed. Two early applications proposed using the fuzzy geometric mean to compute marginal evaluations [41, 283]. Deng [79] uses fuzzy arithmetic means to represent marginal evaluations, forming crisp evaluations by taking a specific  $\alpha$ -cut of each fuzzy mean and forming a weighted average of the left and right points of the resulting interval. These are then standardised and evaluated by comparing them to positive and negative ideal solutions. Mikhailov et al. [191, 192, 193] uses a preference programming formulation to maximise both the consistency of judgements and the possibility i.e. membership values, of those judgements. Leung and Cao [169] define a different fuzzy version of consistency based on specifying a tolerance level for inconsistent ratio judgements, and maximise membership values subject to constraints on this consistency. Yu et al. [297] use a multi-objective linear programming approach to maximise membership values and minimise inconsistency, requiring an additional trade-off parameter. Xu and Chen [294, 293] also minimise the inconsistency of assessed judgements, but for additive preference relations rather than the multiplicative relations common in AHP. Chang [50] proposes a method for triangular fuzzy evaluation called ‘extent analysis’, which first uses the fuzzy arithmetic mean to represent marginal evaluations and then performs pairwise comparisons by computing the possibility that one fuzzy mean is greater than another. For two fuzzy numbers  $A$  and  $A^*$ , the possibility that  $A > A^*$  is given by 1 if the point  $x$  at which  $\mu_A(x) = 1$  is greater than or equal to the point  $y$  at which  $\mu_{A^*}(y) = 1$ , or by the height at which the membership functions intersect otherwise. A final crisp evaluation for  $A$  is computed by taking the conjunction i.e. minimum, of the possibilities that  $A$  is greater than each of the other fuzzy marginal evaluations, and then standardising these to sum to one. Once crisp evaluations are obtained, aggregation is straightforward – Chang [50] uses a simple additive model. The extent analysis approach to fuzzy AHP is currently the most popular approach – by some way – for real-world applications of fuzzy AHP, with some 20 papers published since 2002. However, Wang et al. [277] have recently convincingly argued that the weights obtained using extent analysis model do not represent the relative importance of criteria or alternatives because they confuse evidence of strength with the measurement of that strength.

## Models using comparisons to ideal solutions

Another class of models that have been relatively popular targets for fuzzification are those that evaluate alternatives by comparing them to an ideal and/or anti-ideal solution and choose the alternative that is in some sense closest to the ideal and farthest from the anti-ideal. Most of these models make use of the TOPSIS method proposed by Hwang and Yoon [128]. The TOPSIS method begins by defining two hypothetical alternatives: an ideal solution consisting of the maximum weighted evaluations across all alternatives on each attribute, and an anti-ideal solution consisting of the minimum evaluations. Euclidean distances between each alternative and the ideal and anti-ideal solutions are computed and alternatives are evaluated based on its distance to the ideal solution expressed as a proportion of the sum of the two distances. All that is required in order to use fuzzy input values are computations of fuzzy maxima (for the ideal solution), minima (for the anti-ideal solution), and distances. All of these are standard fuzzy operations.

Differences between the fuzzy TOPSIS methods relate primarily to when in the decision process the fuzzy information is condensed into crisp evaluations. Tsaur et al. [255] do this right at the beginning, representing fuzzy evaluations by their centroids and calculating the crisp distances accordingly. Chu and Lin [68] use fuzzy multiplication to obtain fuzzy weighted ratings but then convert these into crisp values using the method of mean removals [143]. Chen [53] evaluates the ‘Euclidean’ distance between two triangular fuzzy numbers  $\tilde{A} = (l_1, m_1, u_1)$  and  $\tilde{B} = (l_2, m_2, u_2)$  by  $d(\tilde{A}, \tilde{B}) = \sqrt{(l_1 - l_2)^2 + (m_1 - m_2)^2 + (u_1 - u_2)^2}$ , following which the deterministic TOPSIS method can be applied to these crisp distances. Ashtiani et al. [12] use much the same approach, but apply it to interval-valued fuzzy numbers i.e. a fuzzy number that is defined by two membership functions, an ‘upper’ and ‘lower’ membership function ( $\mu_A^U(x)$  and  $\mu_A^L(x)$  respectively, with  $\mu_A^U(x) \geq \mu_A^L(x)$ ). Separate computations are carried out using the upper and lower membership functions and at the final step a simple average is taken of the relative idealities using  $\mu_A^U(x)$  and  $\mu_A^L(x)$ . Whether using double membership functions is a necessary or even useful extension for decision aid seems another matter. Triantaphyllou and Lin [254] use fuzzy arithmetic operations at each step, so that the result is a fuzzy relative ideality for each alternative that must be ranked using one of the fuzzy ranking methods. Their method, contrary to what is claimed in [12], preserves fuzziness right up to the end of the decision process (barring the ranking stage), but Wang and Elhag [275] find that this results in the supports of the fuzzy relative idealities being overexaggerated. Their method uses a series of  $\alpha$ -cuts, calculating the relative ideality of each alternative at each  $\alpha$ -cut. The relative idealities are each interval numbers whose lower and upper bounds can be found using a simple fractional programming model and collected across  $\alpha$ -cuts.

### 2.5.2 Methods for ranking fuzzy numbers

As the evaluations arising from the aggregation in the previous sections are usually fuzzy numbers themselves, the ranking of evaluations is often non-trivial and some kind of ranking procedure is needed. The lack of an axiomatic basis for carrying out such a ranking means that there is no clear guide as to how this ranking should be done, and procedures are most often judged upon (a) how consistent their rankings are with human intuition, (b) how often and in what cases they are unable to discriminate between two different fuzzy numbers, and (c) how difficult they are to interpret and compute. All of these criteria are themselves fraught with subjectivity, and unfortunately it is possible, for nearly all ranking methods, to construct specific instances where (a) or (b) are poorly fulfilled. Research in this area often takes this form, presenting an example where a previous method gives an inconsistent rank order or is unable to distinguish two ‘intuitively’ differentiable fuzzy numbers, and then proposing a new procedure that does better in this particular instance (but possibly not in others).

Chen [59] provides a review of some 21 ranking methods up to 1992. A review to 2001 is provided in [272] and [273], and includes 29 methods. Outside of the ranking methods contained in these reviews, I have been able to find more than 20 further proposed ranking methods. A detailed review of all these methods is not attempted here; instead I have simply classified the more recent methods in Table 2.1 according to which of a small number of general ranking approaches they follow: whether centroids are used to defuzzify fuzzy numbers; whether some form of distance or area measurement is used; whether one or more reference (ideal or anti-ideal) fuzzy numbers are used; and whether any kind of weighting or transformation function is applied. In the remainder of the section I use a small selection of ranking methods to illustrate some general comments.

The main problem when attempting to organise fuzzy ranking methods is that many concepts lack a consistent, meaningful behavioural interpretation. Parameters are often given a dubious behavioural description or are ignored completely. Criticism of the lack of formal procedures for eliciting the parameters of fuzzy decision models is not new (e.g. [100, 31]) but remains a problem. As an example, one stream of papers has made use of the centroid  $(\bar{x}, \bar{y})$  of a fuzzy number – the center of gravity of the area represented by the fuzzy set – to build a single index of performance, an approach that goes back to [201]. Cheng [62] uses the distance between the origin and the centroid as a measure of performance (incidentally giving the wrong formula for computing the centroid – pointed out and corrected in [279]). Subsequent methods proposed using the signed rather than Euclidean distance [1], or the rectangular area formed between the centroid and the origin [69]. It is not clear how these differences would be explained to a decision maker, or

Authors	Year	Ref.	Uses centroid	Uses distance	Uses area	Uses ref. pt.	Uses weighting
Chen and Klein	1997	[54]			y	y	
Cheng	1998	[62]	y	y			
Cheng	1998	[62]					
Raj and Kumar	1998	[220]		y		y	
Lee-Kwang and Lee	1999	[168]			y	y	
Raj and Kumar	1999	[221]		y		y	
Yao and Wu	2000	[296]			y		
Chen and Lu	2001	[56]			y		y
Lee and Li	2001	[167]			y	y	
Mattarazzo and Munda	2001	[189]			y		
Mattarazzo and Munda	2001	[189]					
Yager et al.	2001	[295]			y		
Chu and Tsao	2002	[69]	y		y		
Tran and Duckstein	2002	[253]			y		y
Chen and Chen	2003	[57]	y				y
Deng and Liu	2005	[80]	y			y	y
Liu and Han	2005	[178]	y				y
Abbasbandy and Asady	2006	[1]	y	y			
Deng et al.	2006	[81]			y		
Sun and Wu	2006	[248]					
Asady and Zendehnam	2007	[11]			y		
Chen and Chen	2007	[58]	y			y	y
Chen and Tang	2008	[55]			y		
Liu, Wu and Liu	2008	[177]					
Wang and Lee	2008	[280]	y		y		
Wang et al.	2008	[281]			y	y	
Abbasbandy and Hajjari	2009	[2]			y		y
Chen and Chen	2009	[60]	y				y
Chen and Wang	2009	[61]			y	y	y

Table 2.1: A guide to recent fuzzy ranking methods



whether the selection of one method over another can be informed by any behavioural information. Liu [178] includes the possibility of using weighting functions on  $x$  and  $\mu(x)$  in calculating the centroid. These functions are interpreted as a form of “optimism” and “confidential attitude” respectively, although without any discussion of their behavioural validity beyond a statement that they were “easily understood” and “commonly used in daily decision making”. This seems doubtful. Another method proposes that only  $\bar{x}$  be used to rank fuzzy numbers, with  $\bar{y}$  being reserved for breaking any ties involving the  $\bar{x}$  [280]. Others modify the  $y$  co-ordinate of the centroid to incorporate the standard deviation of the fuzzy number, creating a modified centroid point  $(\bar{x}, \bar{y}^*)$ , but again different authors have modified the centroid in different ways. Some use the standard deviation as a multiplier [58]; some use it as exponent [57, 80]. Again it is difficult to make any meaningful comparison of these approaches. The modified centroid has been used to create a new performance index  $(\bar{x} + \bar{y}^*)$  [57]. This is sometimes used to compute distances to ideal and anti-ideal solutions [80], and sometimes to compute distances to a local minimum (the point  $(\bar{x}_{min}, 0)$ , where  $\bar{x}_{min}$  is the minimum  $\bar{x}$  over all fuzzy numbers to be ranked) [58]. Other authors have simply excluded  $\bar{y}$  and based their index on  $\bar{x}$  and the standard deviation of the fuzzy number only [60].

A related problem is that similarities (and even equivalences) between ranking methods are often masked by the use of different terminology. Two methods [296, 56] independently propose ranking two fuzzy numbers by using the sum of the area between the two left-side membership functions and the area between the two right-side membership functions. The method in [296] is “essentially the same” (in the authors’ own words) as the earlier area compensation method in [99], but the former is phrased in terms of “signed distance”. When this “signed distance” is integrated over all possible  $\alpha$ -cuts it becomes the area in [99]. Fortemps and Roubens [99] prove that the ranking provided by their method corresponds to that provided by the total integral value – the sum of the areas under the inverse functions of the left-side and right-side membership functions as used in [176]. Another method [11] uses an unweighted sum of the same two areas making up the total integral value and is thus equivalent to a special case of [176], although the authors describe their approach in different terms, calling it the “nearest point” to a fuzzy number.

## 2.6 Decision analysis with scenarios

One way of addressing uncertainty about what is going to happen in the external environment is to construct a number of ‘stories’ which describe possible ways in which the future might unfold. Each of these possible futures is known as a ‘scenario’. The most fundamental requirement of a scenario is that it is internally consistent – that is, that it does not contain elements that cannot co-exist with one another, although it does not

imply that a description of a scenario must be complete.

### 2.6.1 Scenarios in scenario planning

The use of scenarios for strategic planning was developed as the more-or-less qualitative technique of scenario planning (e.g. [267, 266, 260]), which also gave the term ‘scenario’ a more specific meaning. In the scenario planning view, a scenario should be constructed to (a) be plausible in the sense that its evolution from the present can be traced in the form of a logical ‘story’, (b) be relevant and useful for the decision makers, in the sense that it provides a comprehensive framework for evaluating and developing future strategy, and provides novel insights into the decision problem. The ideal number of scenarios for achieving a balance between depth and breadth of exploration is said to be between two and four [260]. The scenarios developed under scenario planning are meant to be fairly extreme but still plausible, these being viewed as more likely to generate unusual insights into the problem and also in some sense providing a wider coverage of possible futures, something scenario planners refer to as “bounding the future” [229] (although there will always be more extreme scenarios). They are not necessarily complete descriptions of a single possible future, and the likelihood of occurrence, beyond the existence of plausibility, is not generally a concern when constructing scenarios for scenario planning. Scenarios constructed in this way are considered part of ‘future-focused’ thinking [288] in contrast to approaches which consider uncertainty in terms of what outcomes are possible for a given set of alternatives (an example of ‘alternative-focused’ thinking [145]). Wright and Goodwin [288] argue that scenarios generated using the latter approach are more likely to suffer from well-known heuristics like anchoring and insufficient adjustment and inertia, and are thus less likely to confront decision makers with challenging or surprising views of the future – an important goal for scenario planning.

The implications of different scenario construction for forecasting and general strategic decision making are further explored in [42] and [119]. The latter paper in particular highlights some of the difficulties in evaluating what constitutes a ‘good’ scenario planning intervention. Reported case studies represent only a small and probably positively-biased sample and measuring long-term strategic impact is a particularly difficult task. Different practitioners emphasise different aspects of scenario planning, particularly when it comes to issues around prediction, the use of probabilities, and creativity. What emerges in [119] is a set of several mechanisms by which scenario planning might be beneficial as a tool for decision making rather than any definitive consensus over whether and why this is so.

In the scenario planning view, the constructed set of scenarios are useful in their own right in getting a group of decision makers to express their views on uncertainty and agree on a small number of progressions to future states of the world. It may be possible to

evaluate these scenarios in terms of their desirability (introducing a multi-criteria aspect; see [129] for an example), even though the decision makers have no control over which scenario occurs. More commonly though, the constructed scenarios are used to evaluate and develop alternative strategies. It is at this stage that multi-criteria modelling seems applicable, but many advocates of scenario planning prefer to avoid formal quantitative modelling (e.g. [260, 230]) and use informed but informal judgement – some examples can be found in [95, 285, 44]. Nevertheless some work has been done to integrate decision analysis with the use of scenarios, and the use of scenarios (or “state-contingency”) has also received attention in related areas, with Quiggin claiming that for economic problems “in almost every case, uncertainty is best interpreted in a state-contingent framework” [219].

### 2.6.2 Scenarios in multi-criteria analysis

The main objective of a scenario-based model is to evaluate and compare the performances of alternatives in each scenario – given a decision problem, the approach considers that problem separately in each scenario before (possibly but not necessarily explicitly) combining this information to arrive at a final decision. The general approach can be represented using a value tree in which scenarios are placed in the second level of the hierarchy as parents to  $N_s$  structurally similar “within scenario” value trees, where  $N_s$  is the number of scenarios used. This tree would follow the same format as the tree illustrating quantile-based decision making in Figure 2.2, except that scenarios would take the place occupied by quantiles. Although from a mathematical perspective scenarios can also be included at lower levels of the objectives hierarchy (e.g. as done with quantiles in Figure 2.1), this is considered contrary to the philosophy of scenario planning and would generally not be useful. The general formulation of a multi-attribute scenario model evaluates alternative  $a_i$  by

$$U_i^{(\text{scen})} = \sum_{k=1}^{N_s} \sum_{j=1}^J w_{jk} u_{jk}(z_{ij}^{(s_k)}) \quad (2.7)$$

where  $s_k$  refers to a specific scenario,  $z_{ij}^{(s_k)}$  is the evaluation of alternative  $a_i$  on attribute  $c_j$  in scenario  $s_k$ , and  $w_{jk}$  and  $u_{jk}$  are respectively the weight and marginal utility function associated with attribute  $c_j$  under the assumption of scenario  $s_k$ .

One of the most important issues in integrating the use of scenarios with multi-criteria analysis is how (or whether) to compare and aggregate results in different scenarios. The earliest multi-criteria scenario model (presented in Chapter 14 of [108] and is described in slightly more detail in [109]) applied a multi-attribute value model within each of three scenarios but made no attempt to aggregate the evaluations over scenarios to arrive at a final global evaluation. While this is in line with the philosophy of scenario planning,

which has strong ‘robustness’ views (see [260]), it does not seem unreasonable to attempt to aggregate over scenarios in some way. Applications of scenario planning to the analytic hierarchy process [10, 184, 170, 171] and goal programming [153] have aggregated results over scenarios using some form of relative likelihood to weigh the performance in the different scenarios. Performance has also been aggregated over scenarios using weights interpreted as swing weights for the performance under a particular scenario, in the value-function [244] and goal programming [90] approaches. Kalu [140], on the other hand, solves a lexicographic goal program under various scenarios representing different combinations of operating parameters but stresses the impossibility of probability assignment. The optimal decision vector is thus given separately for each scenario.

Another important consideration is how preference information is to be assessed in the scenario model. Stewart [246] gives two general scenario-based models that are mathematically equivalent but differ practically in that their inputs require the specification of different trade-offs and hence use different assessment procedures. The first model considers combinations of alternatives and scenarios as  $IN_s$  distinct outcomes or ‘meta-alternatives’ to be evaluated in terms of the  $J$  attributes. A marginal preference model (whether this uses a value function, outranking, goal programming, or other approach) is defined across all  $IN_s$  outcomes for each attribute. The result is an  $I \times N_s$  table giving the aggregate performance of alternative  $a_i$  under scenario  $s_k$ . This is the approach followed in [108] and [109], although without a final aggregation over scenarios. If performance is aggregated over scenarios, the evaluation of  $a_i$  is given by

$$U_i^{(\text{scen})} = \sum_{k=1}^{N_s} \left[ w_{s_k} \sum_{j=1}^J w_j u_j(z_{ij}^{(s_k)}) \right] \quad (2.8)$$

where  $w_{s_k}$  is the weight associated with scenario  $s_k$ . If alternatives are not aggregated over scenarios the second summation over  $k$  is simply ignored. As for the quantile weights discussed earlier, the practical interpretation and assessment of the scenario weights  $w_{s_k}$  has not been fully resolved. Stewart [246] argues that the  $w_{s_k}$  should not be equated with scenario “probabilities” (because the set of scenarios does not constitute a complete probability space), nor with scenario “likelihoods” (because scenarios are incomplete descriptions, they cannot in general be expected to represent the same dimensions in probability space). Rather, they should be interpreted as relative (“swing”) weights on performance in different scenarios. This is theoretically permissible, since the  $\sum_{j=1}^J w_j u_j(z_{ij}^{(s_k)})$  values constitute an interval preference scale, but as Stewart notes “it may be difficult to elicit appropriate values for the scenario weights” [246].

The second model considers combinations of scenarios and attributes as  $N_s J$  distinct ‘meta-attributes’, and evaluates the  $I$  alternatives in terms of each of these meta-attributes.

This means that a marginal preference model is constructed for each of the  $N_s J$  meta-attributes, following which performance would be aggregated over all meta-attributes (possibly first within each scenario and then over scenarios, if this is desired). The evaluation of  $a_i$  would now be given by (2.7).

There are a number of issues raised by the two scenario-based models. The first is that it is only in the second model that preference information (trade-offs between attributes and the strength of preference for incremental performance changes on any one attribute) are allowed to vary between scenarios. In one of the few applications of scenario-based MCDA, Montibeller et al. [197] found that in one of their two applications progress was only possible once importance weights were allowed to vary across scenarios – and hence the second model evaluating alternatives over scenario-attribute pairs was used. In the other case one alternative was dominant and this was not necessary. In Korhonen [153] aspirations and weights may be defined separately for individual scenarios. Parnell et al. [211] also explicitly define different importance weights in different scenarios. In other applications preferences are held constant over scenarios [109]. It remains an open question how often the detailed qualitative information gathered during the construction of the scenarios might cause scenario-dependent preferences, or at least an awareness of those preferences.

The practical assessment of preference information will also depend on which model is employed. In the first model, each attribute is defined over a range covering all  $IN_s$  meta-alternatives and weights can therefore be assessed as for a deterministic multi-attribute problem, except that the joint consideration of alternatives and scenarios may prove difficult for decision makers (the case in [197]). The second model requires an importance weight  $w_{jk}$  for each scenario-attribute combination. These can be assessed directly using comparisons between performance levels on different meta-attributes, although this may be impractical for large numbers of attributes or scenarios. A second approach is to first establish the relative weight of each attribute  $c_j$  under the assumption of a common scenario  $s_k$ , denoted  $w_{j|k}$  and standardised to sum to one within each scenario, and then assess the scenario weight  $w_{s_k}$  associated with each scenario  $s_k$  by using trade-offs between pairs of scenarios on a common attribute. The joint weights could then be inferred from the two sets of ratios generated by the previous steps i.e.  $w_{jk} = w_{s_k} w_{j|k}$ .

It is worth mentioning that some multi-attribute scenario applications [155, 129] include elements that are under the control of the decision maker and for which guidance is being sought (what would usually be termed the alternatives) in their constructed scenarios. There are thus no ‘alternatives’ as such: it is the scenarios that are evaluated over attributes, and a preference ordering over scenarios is constructed. Situations in which a

formulation such as this would be more appropriate than the conventional evaluation of alternatives over a set of constructed scenarios would seem to be a rarity<sup>6</sup>, and limited to cases in which alternatives and future events are so interconnected that only one or two alternatives are possible in each scenario. This formulation can be expressed in terms of the first model, except that the ‘meta-alternatives’ are directly constructed through the scenario planning process and it would make no sense to aggregate over scenarios.

University of Cape Town

---

<sup>6</sup>Some authors go further than this, implying that such a formulation goes against scenario planning principles. Harries [119], for example, includes the independence of scenarios and actions as a characteristic of scenario-based decision making.

## Chapter 3

# A simulation study investigating the effect of uncertainty representation on decision making

### 3.1 Introduction

The previous chapter reviewed the many models that exist for multi-criteria decision analysis (MCDA) under conditions of uncertainty. The remainder of the thesis takes a number of these ‘simplified’ decision models and evaluates their ability to replicate the results obtained from a MAUT model, with the aim of providing some guidance to practitioners about the types of simplified models that are being used for uncertain decision making. This chapter performs part of that evaluation using a simulation experiment. As mentioned in Chapter 1, in using simulation it must be acknowledged that while simulations allow one to assess how the simplification of MAUT models might impact results, they cannot evaluate other critical issues like ease-of-use or ability to generate insight. Simulation results alone are unable to provide general conclusions on the viability of different methods, but they provide a starting point for doing so by identifying the potential trade-offs in accuracy that are implied when using a simplified model. Ultimately accuracy – the focus of this chapter – must be weighed against other practical factors to determine which decision model may be most appropriate for a problem.

The models tested here represent uncertain attribute evaluations using (a) expected values, (b) expected values and variances, (c) expected values and the probabilities of obtaining performance below a specified cut-off, (d) quantiles, (e) fuzzy numbers, or (f) a small number of ‘scenarios’. Since these methods all summarise aspects of the full probability distribution, they are referred to collectively as ‘simplified’ approaches; models (b) and (c) are sometimes referred to collectively as models using ‘explicit risk attributes’. All models are based upon the principles of value function methods [e.g. 31]; investigating uncertainty for other schools of MCDA is left to future research.

The chapter is organised as follows. Section 3.2 very briefly summarises the models used (more information can be found in Chapter 2). Section 3.3 lists research hypotheses to be tested by the simulation study. Section 3.4 outlines the simulation experiment, and Section 3.5 presents the results. A final section concludes the chapter with some implications for practice.

## 3.2 Summary of simplified decision models used

Consider a decision problem consisting of  $I$  alternatives denoted by  $a_i$ ,  $i \in \{1, \dots, I\}$ , evaluated on  $J$  attributes denoted by  $c_j$ ,  $j \in \{1, \dots, J\}$ . Let  $Z_{ij}$  be a random variable denoting the (stochastic) attribute evaluation of  $a_i$  on  $c_j$ , and  $u_j(Z_{ij})$  be single-attribute utility functions. Then the additive MAUT model [147] evaluates alternatives by their expected utilities:

$$U_i = \sum_{j=1}^J w_j E[u_j(Z_{ij})] \quad (3.1)$$

where  $U_i$  is the expected utility of alternative  $a_i$  and  $w_j$  is an attribute importance weight indicating the relative importance of a one-unit change in attribute  $c_j$  [e.g. 31]. The additive MAUT model requires that preferences are additively independent [147], failing which more complex aggregation forms are required (although [242] has shown that the additive form can closely approximate results obtained under the more complex multiplicative form when additive independence does not hold). This chapter is concerned with further simplifications of the MAUT model. The following is not exhaustive but provides a broad coverage of the types of simplified models that appear in the literature.

### 3.2.1 Models using expected values and explicit risk attributes

The evaluation of  $a_i$  is given by

$$U_i^{(\text{risk})} = \sum_{j=1}^J \left[ w_j u_j(E[Z_{ij}]) - w_{ij}^R R_{ij} \right] \quad (3.2)$$

where  $R_{ij}$  is a measure of the ‘risk’ of  $Z_{ij}$  and  $w_{ij}^R$  is a weight for  $R_{ij}$ , termed a ‘risk weight’. An expected value model ( $w_{ij}^R = 0$ ) and two explicit risk models are used:  $R_{ij} = \sigma_{ij}^2$  and  $R_{ij} = \Pr[Z_{ij} < \mathcal{L}]$  with  $\mathcal{L}$  a specified cut-off for ‘poor’ performance.

### 3.2.2 Models using quantiles

The evaluation of  $a_i$  is given by

$$U_i^{(\text{quan})} = \sum_{k=1}^{N_q} \left[ w_{qk} \sum_{j=1}^J w_j u_j(z_{ij}^{(qk)}) \right] \quad (3.3)$$



where  $q_k$  refers to a specific quantile,  $z_{ij}^{(q_k)}$  is the  $q_k$ -th quantile of  $Z_{ij}$ ,  $w_{q_k}$  denotes the weight associated with quantile  $q_k$ , and  $N_q$  is the number of quantiles used. Note that here attribute importance weights and utility functions have been kept constant over quantiles (see Section 2.4.3 for details). Two example applications can be found in [245].

### 3.2.3 Models using fuzzy numbers

There are a large number of fuzzy value function approaches, with no general formulation. In one standard approach triangular fuzzy global evaluations are computed using basic fuzzy operations [e.g. 85] as

$$\tilde{U}_i = \left[ \sum_{j=1}^J w_j u_j(z_{ij}^{(q_1)}), \sum_{j=1}^J w_j u_j(z_{ij}^{(q_2)}), \sum_{j=1}^J w_j u_j(z_{ij}^{(q_3)}) \right] \quad (3.4)$$

where  $z_{ij}^{(q_1)}$ ,  $z_{ij}^{(q_2)}$  and  $z_{ij}^{(q_3)}$  denote some lower, intermediate, and upper quantiles respectively. These evaluations can be ranked by any number of fuzzy ranking procedures – the approach used here is based on left and right dominance [56], which is a generalization of the area compensation approach in [99]. The final (crisp) evaluation of  $a_i$  is given by

$$U_i^{(\text{fuzz})} = \frac{1}{\mathcal{N} + 1} \left[ \Theta \sum_{n=0}^{\mathcal{N}} r_{in} + (1 - \Theta) \sum_{n=0}^{\mathcal{N}} l_{in} \right] \quad (3.5)$$

where  $r_{in}$  and  $l_{in}$  are the upper and lower limits of the  $n^{\text{th}}$   $\alpha$ -cut of the fuzzy number  $\tilde{U}_i$  (so-called right- and left-dominance) respectively, and  $\Theta \in [0, 1]$  is the weight assigned to right-dominance. The  $\alpha$ -cuts used are given by  $\alpha_{\tilde{U}_i} = n/\mathcal{N}$  for  $n = \{0, 1, \dots, \mathcal{N}\}$ . It can be shown (see Appendix A.1) that this particular fuzzy decision model is equivalent to a quantile model in which the right-dominance weight  $\Theta$  determines the quantile weights<sup>1</sup>.

### 3.2.4 Models using scenarios

The evaluation of  $a_i$  is given by

$$U_i^{(\text{scen})} = \sum_{k=1}^{N_s} \left[ w_{s_k} \sum_{j=1}^J w_j u_j(z_{ij}^{(s_k)}) \right] \quad (3.6)$$

where  $s_k$  refers to a specific scenario,  $z_{ij}^{(s_k)}$  is the evaluation of alternative  $a_i$  on attribute  $c_j$  in scenario  $s_k$ ,  $w_{s_k}$  is the weight associated with scenario  $s_k$ , and  $N_s$  is the number of scenarios used. Note that here attribute importance weights and utility functions have been kept constant over quantiles (see Section 2.6 for details). Applications of multi-attribute scenario models have been reported in [184, 118] and [197].

<sup>1</sup>Although the equivalence is easily shown it is not transparent *a priori*, and was only noted after similar simulation results were observed for the two models. Although it is perhaps unnecessary to show the quantile and fuzzy models results separately (since the equivalence means that any findings will apply to both models), I have chosen to do so and so retain the original structure of the simulation study in the presentation of the results.

### 3.3 Research aims and hypotheses

The main aim of this chapter is exploratory: to evaluate how closely the simplified models in Section 3.2 approximate the results obtained using an additive MAUT model. Nevertheless there are some expectations which can be formalised as hypotheses. In this section I give a brief motivation for each expected result before explicitly stating these as research hypotheses. It is assumed without loss of generality that utility increases in  $Z_{ij}$  and that each marginal utility function  $u_j$  has been scaled to have a minimum of 0 and a maximum of 1. In [91] it was found that a model using expected values was on average more accurate<sup>2</sup> than a variance model using Kirkwood's [150] weights. Worse accuracy is expected from variance models in which only a general appetite or aversion for risk i.e.  $w_{ij}^R = Cw_j$  with  $C$  a constant, is expressed.

**Hypothesis 1:** Variance models in which risk weights are fixed multiples of attribute importance weights will be less accurate than an expected value model.

Keefer and Bodily [144] have shown that expected values can be closely approximated using the 5%, 50% and 95% quantiles, so three (or more) quantiles could be used to approximate expected values and apply an expected value model. Better accuracy is expected if quantiles are transformed into utilities before aggregation i.e. from the quantile model.

**Hypothesis 2:** Quantile models will be more accurate than an expected value model.

The accuracy of all models will suffer from assessment errors, but theories of "error cancellation" [e.g. 151] suggest that models that use multiple inputs will be more robust to random errors than those that provide more concise summaries of uncertainty.

**Hypothesis 3:** The robustness of model accuracy to assessment errors will be positively correlated with the number of inputs used to summarise probability distributions.

One of the key determinants of the accuracy of the expected value model is the steepness of the marginal utility functions in the region of the approximation, because any differences between MAUT utilities and their approximations are more heavily penalised there [91]. The same effect is expected in all simplified models.

**Hypothesis 4:** The accuracy of all simplified models will worsen as utility functions become steeper in the region of the attribute domain in which approximations are made.

---

<sup>2</sup>All hypotheses and reported results use "utility loss" [19] to measure accuracy, as defined in Section 3.4.7. Other metrics were also gathered, including the average rank of the alternative selected by a simplified model in the MAUT rank order, the average rank of the MAUT best alternative in a simplified model's rank order, and the rank correlation between the simplified model and MAUT rank orders. These did not provide any additional insights and are not reported.

The accuracy of the expected value model is not materially affected by the number of alternatives or attributes present [91]. The same relative insensitivity to problem size is hypothesised for the other simplified models<sup>3</sup>.

**Hypothesis 5:** Problem size (the number of alternatives and attributes) will not have a material effect on the accuracy of any of the simplified models.

Because there has been little systematic research into the use of scenarios in decision analysis, no hypotheses are made regarding these models. An exploratory analysis of the scenario models remains an important aim of the simulation.

### 3.4 Design of the simulation experiment

Figure 1 shows an outline of a single simulation run. Dashed boxes have been used to indicate those parts of the simulation applied iteratively to each alternative and attribute. Corresponding section numbers are shown to indicate where in the text further details can be found (the chapter number is suppressed).

#### 3.4.1 Generating realizations from $Z_{ij}$

The main difference between this and others' simulations of realizations from  $Z_{ij}$  [e.g. 20, 227] is that I consider each  $Z_{ij}$  to be composed of  $L = 10$  normal distributions<sup>4</sup>, denoted by  $N(\mu_{ij\ell}, \sigma_{ij\ell}^2)$ ,  $\ell \in \{1, \dots, L\}$  where  $\mu$  and  $\sigma^2$  are mean and variance respectively. The index  $\ell$  is referred to as indexing the 'future'  $f_\ell$ .

#### Generating means for the realizations in each future

The means  $\mu_{ij\ell}$  are all independently generated on  $U[0, 1]$ .

#### Treatment of dominated alternatives

A simulation parameter determines whether:

- the  $\mu_{ij\ell}$  are used directly, which allows alternatives to be dominated within one or more futures i.e. have a smaller mean than another alternative on every attribute [e.g. 20, 227], or

---

<sup>3</sup>This depends on utility loss being used to measure accuracy. Measures of accuracy based on ranks e.g. the average rank of the selected alternative in the MAUT rank order, differ significantly with the number of alternatives used, but an increasing rank does not necessarily imply a deterioration in decision quality (because the size of the rank order has also increased).

<sup>4</sup>Most previous studies use a single distribution for each  $Z_{ij}$ . Multiple distributions are used here because they allow a parsimonious simulated application of scenario models. The choice of distribution, as well as the number of distributions to use, is somewhat arbitrary, but the robustness of the conclusions has been tested against these choices. Evaluations were also simulated from the uniform distribution, and different numbers of distributions ( $L \in \{6, 10, 20\}$ ) have also been used, with no qualitative differences in the results.

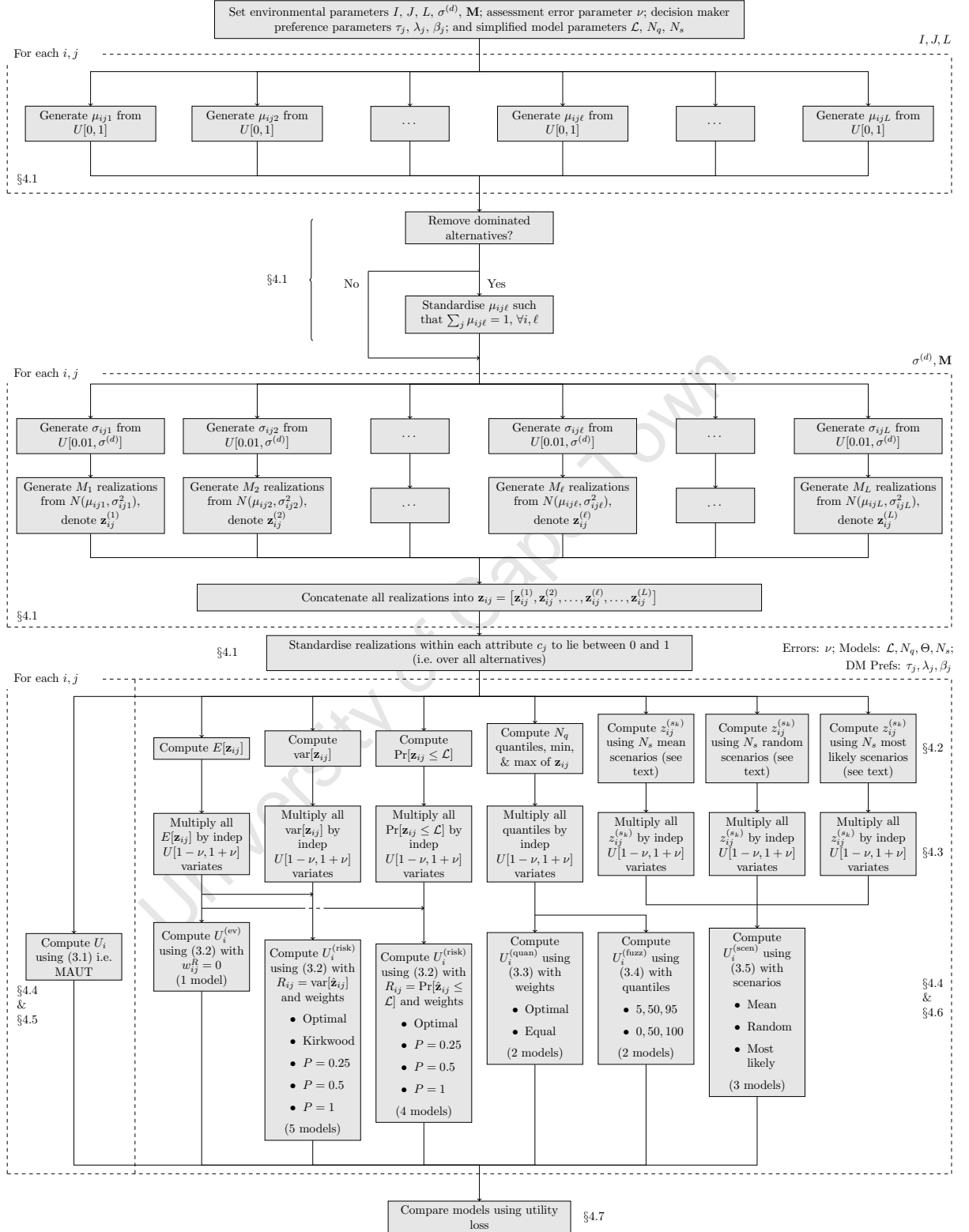


Figure 3.1: Outline of a single simulation run

- alternatives are forced to be Pareto optimal in each future by standardizing within each alternative  $a_i$  so that  $\sum_j \mu_{ij\ell} = 1$  [e.g. 91].

These cases are referred to as “with dominated alternatives” and “without dominated alternatives” respectively.

### Generating unstandardised realizations

Within each future  $f_\ell$ , the simulation

1. Generates a standard deviation  $\sigma_{ij\ell}$  randomly on  $U[0.01, \sigma^{(d)}]$ , where  $\sigma^{(d)}$  is a parameter of the simulation.
2. Sets the number of realizations  $M_\ell$  to be generated for future  $f_\ell$ . A total of  $K = 400$  realizations<sup>5</sup> for each  $Z_{ij}$  is used (i.e. over all futures), and these are distributed over futures either “uniformly” ( $\mathbf{M} = [40, 40, \dots, 40]$ ) or “non-uniformly” ( $\mathbf{M} = [60, 60, 60, 60, 40, 40, 20, 20, 20, 20]$ ). In the latter case some futures are more likely to occur than others.  $\mathbf{M}$  is a parameter of the simulation.
3. Generates  $M_\ell$  independent realizations from  $N(\mu_{ij\ell}, \sigma_{ij\ell}^2)$ . The  $1 \times M_\ell$  vector of realizations generated in future  $f_\ell$  is denoted  $\mathbf{z}_{ij}^{(\ell)}$ .

Once realizations have been generated for each future, these are concatenated into a single  $1 \times K$  vector containing all the realizations for  $Z_{ij}$  i.e.  $\mathbf{z}_{ij} = [\mathbf{z}_{ij}^{(1)}, \mathbf{z}_{ij}^{(2)}, \dots, \mathbf{z}_{ij}^{(\ell)}, \dots, \mathbf{z}_{ij}^{(L)}]$ .

### Standardizing the realizations

Realizations are scaled so that the largest realization on each attribute (over all alternatives) is one and the smallest is zero.

#### 3.4.2 Generating inputs to the simplified models

Each simplified model uses a different summary of the realizations in  $\mathbf{z}_{ij}$ :

##### Expected value model

Uses the empirical mean of  $\mathbf{z}_{ij}$ ,  $E[\mathbf{z}_{ij}]$ .

##### Explicit risk models

Two explicit risk models are simulated (both of these also use the expected values  $E[\mathbf{z}_{ij}]$ ):

1. *Variance model*: uses empirical variances  $\text{var}[\mathbf{z}_{ij}]$  to measure risk
2. *Probability of poor performance model*: uses the proportion of realizations in  $\mathbf{z}_{ij}$  that fall below a cut-off  $\mathcal{L}$ , a parameter of the simulation, to measure risk.

---

<sup>5</sup>Conclusions are insensitive to this choice

## Quantile models

Uses empirical quantiles of the  $\mathbf{z}_{ij}$ . The number of quantiles  $N_q$  is a parameter of the simulation. If  $N_q = 3$ , the 5%, 50%, and 95% quantiles are used; if  $N_q = 5$ , the lower and upper quartiles are added. For the fuzzy model, a triple consisting of the 5%, 50% and 95% quantiles is used i.e.  $\mathbf{q} = (0.05, 0.5, 0.95)$ , allowing a comparison with the quantile models; a triple in which the 5% and 95% quantiles are replaced with the minima and maxima respectively is also used i.e.  $\mathbf{q} = (0, 0.5, 1)$ , these being the extremes most often used in fuzzy analysis (e.g. [167, 96]).

## Scenario models

Inputs to the simulated scenario models are generated in a two-step process. First, a random sample is drawn from the set of futures  $\{f_1, f_2, \dots, f_L\}$  (the same random sample is used for all alternatives and attributes); then, realizations in each of the sampled futures are summarised, for each alternative and attribute. Scenario models differ with respect to how the random sampling and summarization are performed. The following three models are used:

- *‘Mean scenario’ model*
  1. Randomly draws a sample of size  $N_s$  from  $\{f_1, f_2, \dots, f_L\}$  without replacement. Each future  $f_\ell$  has an equal probability of selection.
  2. Summarises performance of  $a_i$  on  $c_j$  using the mean  $\mu_{ij\ell}$  in each of the  $N_s$  selected futures.
- *‘Random scenario’ model*
  1. Randomly draws a sample of size  $N_s$  from  $\{f_1, f_2, \dots, f_L\}$  without replacement. Each future  $f_\ell$  has an equal probability of selection.
  2. Summarises performance of  $a_i$  on  $c_j$  by randomly selecting one realization from  $\mathbf{z}_{ij}^{(\ell)}$  in each of the  $N_s$  selected futures.
- *‘Most likely scenario’ model*
  1. Randomly draws a sample of size  $N_s$  from  $\{f_1, f_2, \dots, f_L\}$  with replacement. Each future  $f_\ell$  is selected with probability proportional to  $M_\ell$ .
  2. Summarises performance of  $a_i$  on  $c_j$  by randomly selecting one realization from  $\mathbf{z}_{ij}^{(\ell)}$  in each of the  $N_s$  selected futures.

The proportion of futures selected i.e.  $N_s/L$ , is a simulation parameter termed the ‘coverage’ provided by a scenario model. Although attribute generation is somewhat biased in favour of a ‘mean scenario’ model using all  $L$  futures, the view taken here

is that the coverage parameter captures (in an idealised way) the scenario planning aim of constructing scenarios that “bound the future” [e.g. 229]; a scenario model with 100% coverage is practically unrealistic but useful in giving an upper bound on accuracy. More realistic scenario models with less coverage (50% and 30%) are also simulated, and sensitivity to coverage and construction method are important results.

### 3.4.3 Generating errors in the assessment of uncertainty information

Assessment errors are simulated by multiplying all inputs to the simplified models (expected values, variances, probabilities of poor performance, quantiles, and realizations within selected futures) by independent and randomly generated realizations on  $U[1 - \nu, 1 + \nu]$ , with  $\nu$  a simulation parameter. Final assessments are denoted using the ‘hat’ symbol e.g. expected values  $\hat{E}[\mathbf{z}_{ij}]$ .

### 3.4.4 Generating preference structures

The simulated marginal utility functions exhibit diminishing sensitivity and loss aversion i.e. are convex below a reference level, concave above it, and steeper below the reference level, *a la* prospect theory [139]. Each marginal utility function is described by four parameters: the reference level,  $\tau_j$ , the value of the utility function at the reference level,  $\lambda_j$ , and the curvature of the utility function below and above the reference level,  $\alpha_j$  and  $\beta_j$  respectively, using the standardised exponential form

$$u_j(x) = \begin{cases} \frac{\lambda_j(e^{\alpha_j x} - 1)}{e^{\alpha_j \tau_j} - 1} & \text{for } 0 \leq x \leq \tau_j \\ \lambda_j + \frac{(1 - \lambda_j)(1 - e^{-\beta_j(x - \tau_j)})}{1 - e^{-\beta_j(1 - \tau_j)}} & \text{for } \tau_j < x \leq 1 \end{cases} \quad (3.7)$$

The same approach was used in [243] and [91]; a diverse set of preferences may be simulated by adjusting  $\tau_j$ ,  $\lambda_j$  and  $\beta_j$  ( $\alpha_j$  is set to  $\beta_j + U[0, 2]$ ). Attribute importance weights are generated to be uniformly distributed with a minimum normalised value of  $1/2J$ , following [43].

### 3.4.5 Simulating the application of an additive MAUT model

The expected utility of  $a_i$  is given by (3.1), where (as in other models) the vector of realizations  $\mathbf{z}_{ij}$  is used in place of the random variable  $Z_{ij}$ . Note that all probability information is taken into account in the generation of  $\mathbf{z}_{ij}$ .

### 3.4.6 Simulating the application of the simplified models

#### Expected value model

The evaluation of  $a_i$  is given by (3.2) with all  $w_{ij}^R = 0$ .

## Explicit risk models

The evaluation of  $a_i$  is given by (3.2). Risk weights are simulated using three approaches:

1. *Using a fixed risk multiplier:* risk weights are set so that the average contribution made by the risk components over all alternatives is a proportion  $P$  (termed the ‘fixed risk multiplier’) of the average contribution made by the value components. ‘Fixed risk weights’ are given by  $w_{ij}^R = Pw_j u_j(E_i[\hat{E}[\mathbf{z}_{ij}]])/E_i[\hat{R}_{ij}] \quad \forall i$ , where  $E_i$  denotes that expectations are taken over all alternatives and  $P \in \{0.25, 0.5, 1\}$ .
2. *Using an optimal risk multiplier:* risk weights are again a constant proportion of the  $w_j$  but with the constant chosen to minimise the rank of the selected alternative in the MAUT rank order. This is implemented with an integer program and the results termed ‘optimal risk weights’. In practice one would not know the MAUT rank order and thus the approach is not practically feasible, but it provides an upper limit on accuracy with a single risk multiplier.

When using variances a further approach is included:

3. *Using Kirkwood’s weights:* Following [150] I use risk weights  $w_{ij}^R = (-1/2)w_j u_j''(\hat{E}[\mathbf{z}_{ij}])$ .

## Quantile models

The evaluation of  $a_i$  is given by (3.3). Two approaches for generating quantile weights are used. The first uses equal weights for each quantile i.e.  $w_{q_k} = 1/N_q$ . The second computes optimal quantile weights using a similar integer program to the one above, except that three (or five) weights can vary and quantile weights are constrained to sum to one.

## Fuzzy models

The evaluation of  $a_i$  is given by (3.4) and (3.5). The simulation uses  $\mathcal{N} = 2$  but results are independent of this value. The weight  $\Theta$  has been interpreted as an index of ‘optimism’ in [56] and is a parameter of the simulation,  $\Theta \in \{0.25, 0.5, 0.75\}$ <sup>6</sup>.

## Scenario models

The evaluation of  $a_i$  is given by (3.6). All scenarios are equally weighted<sup>7</sup> i.e.  $w_{s_k} = 1/N_s$ .

<sup>6</sup>These correspond to quantile weights (for the 5%, 50%, and 95% quantiles respectively) of (0.375, 0.5, 0.125) for  $\Theta = 0.25$ , (0.25, 0.5, 0.25) for  $\Theta = 0.5$ , and (0.125, 0.5, 0.375) for  $\Theta = 0.75$  (see Appendix A.1 for details)

<sup>7</sup>I also tried weighting scenarios by their relative likelihoods i.e.  $M_\ell/K$ , finding that this only improved accuracy substantively when one scenario was overwhelmingly more likely to occur (e.g. 80% of all realizations generated from the same future). For brevity attention is restricted to the case of equal scenario weights.



### 3.4.7 Comparing results of the MAUT and simplified models

The accuracy of each simplified model is evaluated using utility loss [19]. Utility loss is defined as  $UL = (U_{i^*} - U_{i^{sel}})/(U_{i^*} - U_{i_*})$  where  $U_{i^*}$  and  $U_{i_*}$  are the utilities (according to MAUT) of the best and worst alternatives in the MAUT rank order respectively, and  $U_{i^{sel}}$  is the utility (according to MAUT) of the alternative selected by a simplified model. Note that minimizing the rank of the selected alternative in the MAUT rank order (as done by the integer programming formulations above) is equivalent to minimizing utility loss.

### 3.4.8 Parameter values used in the simulations

Table 3.1 provides the parameter values used to simulate the decision problems. The effect of problem size is investigated using  $I = 9$  or  $29$  alternatives (so that a random choice of alternative would appear on average 5th or 15th in the MAUT rank order respectively) with  $J = 10$  or  $20$  attributes. The four combinations allow for an independent investigation of alternatives and attributes in decision environments located between ‘fairly small’ and ‘fairly big’ ([20] and [227] also include smaller problems but otherwise use similar values).

Parameter values for  $\sigma^{(d)}$  were chosen by varying these until realizations in different futures could not be distinguished and appeared sufficiently uncertain (in the ‘low’ and ‘high’ variability conditions the average difference between the 5th and 95th percentile of  $\mathbf{z}_{ij}$  is 0.42 and 0.58 respectively). The assessment error parameter  $\nu$  is key and so is varied at four levels from 0% (error-free) to 30% (severely flawed assessment).

Parameter values for  $\tau_j$  and  $\lambda_j$  can give preferences that are mostly convex (e.g.  $\tau_j = 0.85, \lambda_j = 0.15$ ), mostly concave (e.g.  $\tau_j = 0.15, \lambda_j = 0.85$ ), ‘S-shaped’ or linear (e.g.  $\tau_j \approx \lambda_j$ , with  $\alpha_j$  and  $\beta_j$  both large or small respectively). A wide range of preferences can be simulated with relatively few parameters. The same parameter values have been used in [243] and [91].

Table 3.2 shows the values used for the parameters of the simplified models. The fixed risk multiplier  $P$  used in the explicit risk models is varied between 0 (i.e. the expected value model) and 1 using intervals of 0.25. In assessing the probability of performing below a cut-off, the simulation uses two cut-offs ( $\mathcal{L} = 0.05, 0.10$ ) that represent very poor performance (between the 0.5% and 5% quantiles of performance, depending on attribute variability) and one cut-off ( $\mathcal{L} = 0.50$ ) representing mediocre performance (between 40% and 70% quantiles). Three- and five-quantile summaries are selected as standard summary statistics that are regularly used to approximate probability distributions [102] and moments [144]. Quantiles are chosen for the fuzzy decision model to allow for a direct comparison with quantile models, and to reflect the common practice of defining triangular

Parameter	Description	Values
<b>Problem context:</b>		
$I$	number of alternatives	9, 29
$J$	number of attributes	10, 20
$\mathbf{M}$	distribution of realizations over futures	Uniform: [40, 40, 40, ..., 40] Non-uniform: [60, 60, 60, 60, 40, 40, 20, 20, 20, 20]
<b>Attribute evaluations:</b>		
$\sigma^{(d)}$	upper limit for $\sigma_{ijl}$	0.05 or 0.10
<b>Errors in assessments of uncertainty information:</b>		
$\nu$	width of interval for random factor $U[1 - \nu, 1 + \nu]$ generating errors	0, 0.1, 0.2, 0.3
<b>Marginal utility functions:</b>		
$\tau_j$	reference level for $u_j$	$U[0.15, 0.4]$ or $U[0.6, 0.85]$
$\lambda_j$	value of $u_j$ at the reference level	$U[0.15, 0.4]$ or $U[0.6, 0.85]$
$\beta_j$	curvature of $u_j$ above reference level	0, $U[0, 2]$ or $U[0, 5]$

Table 3.1: Parameter values used to simulate hypothetical decision problems

membership functions between an absolute minimum and maximum [167, 96]. The main goal in selecting the number of scenarios is to investigate the effect of omitting futures – the selection of ten, five and three of the original  $L = 10$  futures, giving ‘coverage’ of 100%, 50% and 30% respectively, is simulated. Note that this does not test the general effectiveness of using three, five, or ten futures, even in the limited sense of accurate approximation of MAUT.

Parameter	Description	Values
$P$	fixed risk multiplier	0.25, 0.5, 1
$\mathcal{L}$	poor performance cut-off	0.05, 0.1, 0.5
$N_q$	number of quantiles used	3, 5
$\mathbf{q}$	quantiles used in fuzzy model	(0.05, 0.5, 0.95), (0, 0.5, 1)
$\Theta$	right-dominance weight (‘optimism’)	0.25, 0.5, 0.75
$N_s/L$	coverage	30%, 50%, 100%

Table 3.2: Parameters used to simulate the application of simplified decision models

A resolution V fractional factorial design [e.g. 195, chapter 8] is used, so all main effects are unconfounded with two- and three-factor interactions, and all two-factor interactions are unconfounded with other two-factor interactions. The simulation performs 100 runs for each combination of parameters, giving standard errors of at most 0.003 for mean utility losses in groups formed by combinations of two factors. This is small enough for any differences discussed in the results to be statistically significant at the 1% level.

### 3.5 Results

Figure 3.2 shows the average utility loss of each simplified model under error-free (indicated by unshaded circles) and erroneous assessments (indicated by shaded circles (10% error), squares (20% error) or triangles (30% error)). Within each model type, utility losses are ordered from best to worst according to the error-free values. This allows one to see the smallest/best mean utility losses as well as the range of possible values. For comparative purposes, a random selection policy gives a utility loss of approximately 0.50 whether dominated alternatives are removed or not. The mean utility of the MAUT best (worst) alternative is 0.43 (0.33) when dominated alternatives are removed, and 0.52 (0.32) when they are included. The larger range of utilities when dominated alternatives are included means that utility *losses* are generally smaller in that condition – because dominated alternatives are very rarely selected by any of the simplified models – but conclusions are not affected by the presence of dominated alternatives. The discussion below focuses mainly on results obtained without dominated alternatives.

In Figure 3.2, Hypothesis 1 and 2 are strongly supported and Hypothesis 3 is conditionally supported. The average utility loss using expected values is better than any explicit risk model using fixed risk multipliers (Hypothesis 1), provided assessment errors are not large. When assessment errors are large a model using probabilities of performing below a central quantile can be more accurate than one using expected values. This suggests that an explicit risk attribute may impart some robustness to assessment errors. In general though, it appears that a model using a fixed risk multiplier approximates MAUT relatively poorly. Sensitivity to assessment errors decreases as the risk multiplier is increased because ‘risk’ components are less sensitive to errors than ‘value’ components (if only variances are used, the utility loss varies from 0.40 to 0.45 depending on assessment errors). The utility losses of variance models with large risk multipliers do not improve when dominated alternatives are included (as other models do) because there are no substantial differences in the variances of dominated and non-dominated alternatives. A variance model using Kirkwood’s weights [150] performs better than one using fixed risk multipliers, but it is only in the unrealistic case where the risk multiplier is optimally chosen that an explicit risk model gives consistently better results than expected values alone.

The average utility loss for the equal-weight quantile models over all assessment errors (in the absence of dominated alternatives) is 0.052 and 0.080 using five or three quantiles respectively, significantly lower than the 0.106 obtained using expected values (Hypothesis 2). The better accuracy of the quantile models is partly due to their greater robustness to

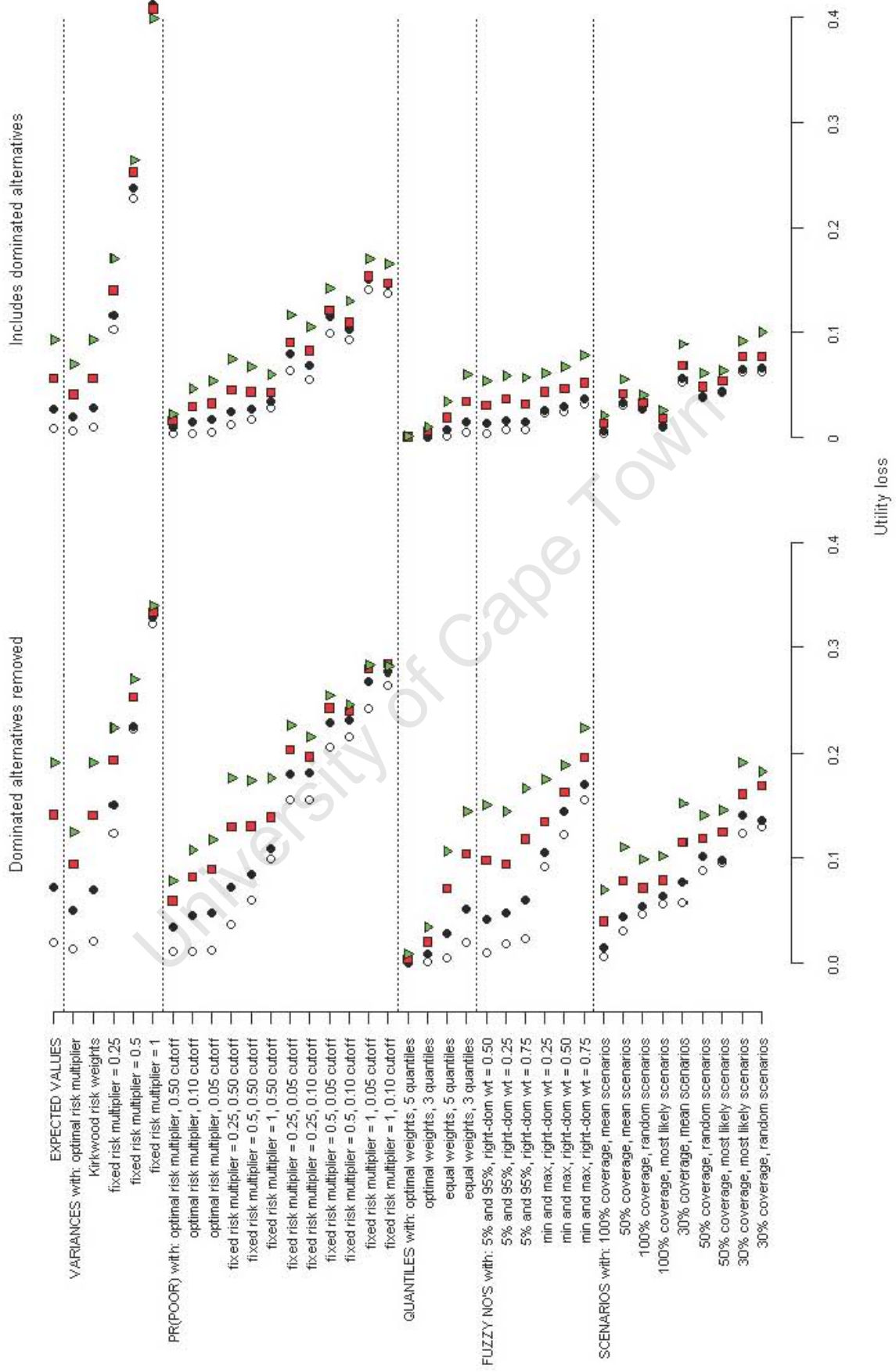


Figure 3.2: Mean utility loss experienced by each simplified model. Unshaded and (three) shaded points show the average utility loss in error-free simulations and with assessment errors of 10%, 20%, and 30%.

assessment error, but even in the absence of errors average utility loss is 0.004 and 0.019 if five or three quantiles are used respectively, and 0.020 using expected values. If quantile weights are chosen optimally, the error-free mean utility loss is less than 0.001 using either five or three quantiles. The average weight allocated to the 5%, 50%, and 95% quantiles is 0.16, 0.65 and 0.19 respectively<sup>8</sup>; this is very close to the weights proposed by [144]: 0.185, 0.630 and 0.185. If five quantiles are used, the average quantile weights are 0.07, 0.12, 0.60, 0.12, and 0.09 for the 5% through 95% quantiles respectively. The increase in accuracy derived from placing more weight on the central quantile is also (as expected) exhibited by the fuzzy models, where the best accuracy is obtained using 5%, 50% and 95% quantiles with  $\Theta = 0.5$  (equivalent to a quantile model using the same quantiles with  $w_{q_1} = 0.25, w_{q_2} = 0.5, \text{ and } w_{q_3} = 0.25$ ). Decreases in accuracy observed when using different weights are only marginal, but large deteriorations are observed if the 5% and 95% quantiles are replaced by minima and maxima respectively.

Figure 3.2 also shows that if one excludes those explicit risk models which perform terribly, there is a clear trend towards increased robustness in the quantile and scenario models, although robustness can also vary widely within model type. Table 3.3 shows the deteriorations in utility loss that occur as a result of different sized assessment errors. The selected models are the best-performing versions of each simplified model (in terms of mean utility loss with no assessment error) that do not make use of optimal weights and have been ranked from the smallest average increase in utility loss over all assessment errors to the largest. Although there is some contrary evidence in that the 3-scenario model is more robust than the 5-quantile model, there is a clear association between the number of inputs used by a simplified model and its robustness to assessment error.

Scenario model accuracy is strongly influenced by both scenario construction and coverage. Substantially better accuracy is obtained if scenarios are constructed using mean values. The relatively poor results obtained when selecting realizations at random from each future highlights the importance of accurately assessing means. If no futures are omitted (100% coverage), then results can be excellent; but if coverage drops to 50% then accuracy (when no assessment errors are made) is worse than if expected values are used. As indicated, the view taken here is that results using 50% and 30% coverage probably provide more appropriate indicators of the practical potential of scenario models. Coverage becomes more important relative to scenario construction when dominated alternatives are present; all models with 100% coverage outperform all those with 50% coverage, which in turn outperform all those with 30% coverage. This is because alternatives can perform terribly

---

<sup>8</sup>These are calculated based on error-free assessments only. When assessment errors are made the simulation chooses optimal quantile weights to compensate for the errors – something real-world decision makers cannot do.

Model	Conditions	Assessment error		
		10%	20%	30%
Scen	10 mean scenarios (100% coverage)	0.008	0.034	0.063
Scen	5 mean scenarios (50% coverage)	0.013	0.048	0.080
Scen	3 mean scenarios (30% coverage)	0.020	0.058	0.095
Quan	5 equally-weighted quantiles	0.024	0.067	0.102
Quan	3 equally-weighted quantiles	0.032	0.085	0.125
Fuzzy	$\mathbf{q} = (0.05, 0.5, 0.95)$ , $\Theta = 0.5$	0.032	0.088	0.141
Pr(Poor)	0.50 cutoff, risk multiplier = 0.25	0.036	0.093	0.140
Var	Kirkwood's weights	0.049	0.120	0.170
EV	–	0.052	0.121	0.172

Table 3.3: Increases in mean utility loss caused by assessment errors. Results are only reported for simulations where dominated alternatives have been removed. Simplified models are represented by their best-performing versions, after excluding models using optimal risk or quantile weights.

in omitted futures. An advantage of scenario models is increased robustness to assessment errors. When assessment errors are made, the mean scenario model with 50% coverage offers similar performance to a three-quantile model using equal quantile weights.

Table 3.4 shows how average utility loss differs over other simulation parameters. The utility function parameters  $\tau_j$  and  $\lambda_j$  are shown jointly to evaluate Hypothesis 4. Accuracy is primarily affected by the shape of the utility functions, and is best when these are predominantly concave and worst when they are predominantly convex. Many of the approximations used in the current simulations occur in the middle-to-upper part of the attribute domain, where convex utility functions are steeper than concave ones (Hypothesis 4). This occurs because any differences are more heavily penalised by a steeper utility function<sup>9</sup>. Further results show that both the approximations made by simplified models and assessment errors cause the deteriorations in accuracy. With no assessment errors, the greatest utility loss occurs (for all simplified models) with highly convex utility functions (high  $\tau_j$ , low  $\lambda_j$ ). The increase in average utility loss caused by the same size assessment error is also greatest when utility functions are highly convex. Table 3.4 also shows that the accuracy of all the simplified models is very nearly constant over the simulated problem sizes<sup>10</sup> (Hypothesis 5). The only other variable exerting a meaningful effect on accuracy is the variability of the attribute evaluations – as evaluations become more variable the

<sup>9</sup>Another explanation is that a sharp threshold at a low reference level implies a utility function that is relatively flat over a large portion of the attribute domain, which reduces the utility loss of selecting a ‘good-but-not-the-best’ alternative. However, the rank of the alternative selected by the simplified models in the MAUT rank order also deteriorates (from an average of 2.0 to 2.4 as utility functions shift from mostly concave to mostly convex), and thus this cannot completely explain the effect.

<sup>10</sup>This depends on utility loss being used to measure accuracy. Measures of accuracy based on ranks e.g. the average rank of the selected alternative in the MAUT rank order, differ significantly with the number of alternatives used, but an increasing rank does not necessarily imply a deterioration in decision quality (because the size of the rank order has also increased).

accuracy of all the simplified models worsens.

### 3.6 Discussion

The simulation experiment reported here takes a number of simplified decision models used in practice and evaluates how close their results are to what might be achieved using MAUT. These results are not intended to be used to conclude a detailed apparatus which prescribes rules for using particular decision models in particular situations. Rather the results suggest a general course of action for practitioners who for reasons of simplicity prefer not to use MAUT. It must be stressed that all findings are limited by the range of cases which have been simulated, as all simulation experiments are. The complexity of the simulation apparatus is largely to ensure that a suitable range of decision problems have been covered (although doubtless there are counterexamples to the findings which could be constructed). Although the apparatus of the simulation experiment may be complex, the conclusions are fairly simple.

The basic message emerging from the simulations is that – for a wide range of simulated decision problems – all of the simplified models are able to produce results that are, on average, close to what would be achieved under MAUT. The best-performing of each of the simplified models have average utility losses of less than 0.04 (where 0 is optimal, and a policy of randomly selecting an alternative has an average utility loss of approximately 0.5). Given the time and effort required to implement a MAUT model, the use of simplified models appears justifiable for many decision problems. The results suggest that avoiding assessment errors in the application of a simplified model is more important than the choice of a particular type of model. This provides numerical support for the notion that analysts have considerable scope to choose the model that they (or the decision makers they are facilitating) are most comfortable with and are least likely to apply poorly. One check that analysts should perform before using a simplified model is to test whether preferences are highly non-linear. The results indicate that the accuracy of all the simplified models deteriorates as preference thresholds become sharper in the region where approximations are made. The main reason for this is that any assessment errors are more heavily punished. Analysts wishing to use simplified models but finding that strong preference thresholds exist should therefore place extra effort in ensuring accurate approximations, for example by using additional consistency checks.

Although the performance of all the simplified models is good in the absence of assessment errors, a quantile model (including those fuzzy decision models which are equivalent to quantile models) performs better than any other model. The final choice of a simplified

Effect	EV	Var	Pr (Poor)	Quantiles			Scenarios: coverage =		
				$N_q = 3$	$N_q = 5$	Fuz	30%	50%	100%
$I$	9	0.10	0.10	0.07	0.05	0.07	0.10	0.06	0.03
	29	0.10	0.10	0.08	0.05	0.07	0.10	0.06	0.03
$J$	10	0.10	0.10	0.07	0.05	0.07	0.10	0.06	0.03
	20	0.10	0.10	0.08	0.05	0.07	0.09	0.06	0.03
$M$	Uniform	0.09	0.09	0.08	0.07	0.05	0.09	0.06	0.03
	Non-unif.	0.10	0.10	0.07	0.07	0.05	0.10	0.06	0.03
$\sigma^{(d)}$	U[0.01,0.05]	0.08	0.09	0.05	0.04	0.05	0.06	0.04	0.02
	U[0.01,0.10]	0.10	0.10	0.08	0.05	0.08	0.11	0.07	0.03
$(\tau, \lambda)$	(low,low)	0.08	0.08	0.05	0.03	0.05	0.09	0.05	0.02
	(low,high)	0.07	0.08	0.04	0.02	0.03	0.07	0.04	0.01
	(high,low)	0.13	0.12	0.12	0.08	0.10	0.13	0.09	0.05
	(high,high)	0.10	0.10	0.08	0.05	0.07	0.10	0.06	0.03
$\beta$	0	0.09	0.09	0.07	0.04	0.06	0.09	0.06	0.03
	U[0,2]	0.10	0.09	0.10	0.05	0.07	0.09	0.06	0.03
	U[0,5]	0.11	0.11	0.10	0.05	0.07	0.11	0.07	0.03

Table 3.4: Average utility losses at different levels of the simulated decision problem parameters. Averages are calculated over all levels of the remaining parameters. Results are only reported for simulations without dominated alternatives. The same models are used as in Table 3.3: the variance model uses Kirkwood weights; the probability of poor performance model uses a 0.50 cutoff and a risk multiplier of 0.25, both quantile models use equal quantile weights; the fuzzy model uses  $\Theta = 0.5$ , and all scenario models use mean scenarios.



model will need to take into account other practical factors. The results only suggest that if analysts wish to conform to MAUT but lack the resources to implement the full model, they should consider a quantile model first. Accuracy is best served by avoiding the use of absolute maxima and minima and by aggregating evaluations over quantiles using the weights proposed by [144]: the median receives roughly two-thirds of the weight and the remainder is shared between the other quantiles. These quantile weights do not provide information about decision maker preferences but rather aim only to approximate a MAUT model. This means that no assessment of quantile weights is required. Quantiles (particularly non-extreme quantiles) can be assessed relatively accurately by decision makers [e.g. 217, 103] and are commonly used in the practical assessment of probability distributions using, for example, the bisection or interval methods [e.g. 238]. Trade-off judgements are also likely to be easier than with other risk measures because quantiles are measured on the same scale as the attribute evaluations. All this suggests a use for the three-quantile model as a preliminary screening tool before a more detailed assessment using MAUT. This could be done to select the most promising alternatives from a larger set, or to assess whether the choice of a preferred alternative is clear enough that a full MAUT analysis is not required. If the bisection (or similar) method is to be used to assess probability distributions then the quantiles assessed for a preliminary model could be used for later assessments too, so that the use of a preliminary model would not add a significant amount of time or effort to the analysis. If a quantile model is to be used on its own, the accuracy results as well as the relative inaccuracies reported when assessing extreme quantiles [8] suggest that upper and lower quartiles be assessed as well as the usual median and extremes. It may be useful to structure the quantiles into scenario-like arrangements by collecting together all the attribute evaluations at a particular quantile. Although such an arrangement is not necessary, some insight might be gained from allowing decision makers to compare alternatives at their ‘almost-worst-case’, ‘almost-best-case’ and some ‘intermediate’ levels of performance. Some evidence for this conjecture is presented in an application reported in Chapter 5.

Several authors have called for decision analysts to pay greater attention to scenario planning techniques [108, 197], which are well-established in strategic decision making. The simulation results indicate that when a substantial number of futures are omitted a scenario model gives a worse average approximation to MAUT than any other simplified model. The time and cognitive effort involved in constructing scenarios means some omissions are inevitable, in my view. Scenario-based MCDA may possess other advantages – generating insights into uncertainty and novel actions are commonly-cited benefits [230] – but analysts should be aware that a scenario model, even if correctly applied, will lead to an outcome that is more different to MAUT than other simplified models. Sometimes an

analyst may for pragmatic reasons (e.g. familiarity) want to use scenarios but also wish to obtain results that are close to MAUT. The results here suggest that in such cases scenarios should attempt to capture mean performances in as many futures as is practically feasible. This is quite different to the philosophy of scenario planning, which often advocates taking a small number of extreme positions when constructing scenarios [e.g. 230].

For the range of simulated decision problems, explicit risk models performed poorly relative to other simplified models. When no assessment errors are made both the variance model and a model using probabilities of poor performance can lead to poorer approximations of MAUT than a model using expected values only, unless risk weights are chosen optimally. The variance model does particularly poorly when risk weights are assigned as a fixed multiple of the attribute importance weights. Behavioural research suggests that the elicitation and understanding of variance information is difficult [e.g. 102], and the assessment of weights involving variances also seems a difficult prospect. In conjunction with the accuracy results, this suggests that the variance model must make use of the risk weights provided by Kirkwood [150] if it is to be applied in practice.

## Chapter 4

# A choice experiment investigating the effect of uncertainty representation on decision making

### 4.1 Introduction

This chapter describes a choice experiment investigating the effects of using different formats for representing uncertain attribute evaluations on decision making. The experiment involves a series of choices presented to numerate but unfacilitated study participants. Each choice task requires that the subject selects one alternative from a set of five. Subjects are informed that in each problem they will be selecting an investment for a friend who has specific aims on one, two, or three objectives. The performance of each alternative on each objective is not known with certainty but is presented to the decision maker using one of six uncertainty formats: probability distributions; expected values with or without standard deviations; a set of five quantiles; a triangular approximation to the probability distribution (minimum-median-maximum); and a set of three representative “scenarios”. No elicitation of uncertainty information is required from the participants – all information is provided to them (in one of the formats mentioned above, without any assessment errors) as part of the task. Because the choice is to be made on behalf of a friend whose preferences are largely specified in the problem, it is possible to identify a ‘true’ best alternative according to an idealised application of MAUT and evaluate the chosen alternative relative to this alternative using utility loss, as in the previous chapter. Information is also gathered on the difficulty experienced by subjects in making a decision, the mental processes by which decisions are reached, and certain characteristics of the selected alternative.

This study continues the line of enquiry begun in the previous chapter by examining how different formats for representing uncertainty can systematically influence decision making, in terms of the outcomes listed above. Although the study is limited in that it does

not study real prescriptive decision making but rather describes naïve subjects making choices in the absence of any facilitation, the results have some implications for prescriptive decision making practice in that they identify natural tendencies which may need to be overcome in the course of a prescriptive analysis. The chapter is organised as follows. The next section presents the main hypotheses to be tested against the experimental data. Following that, a description of the design and execution of the experiment is given, followed by the results. A final discussion section concludes the chapter.

## 4.2 Research hypotheses

Previous research in behavioural decision making suggests that decision makers may be overwhelmed by a large amount of information such as that presented by the full probability distributions [186], and thus that decision quality (interpreted here as synonymous with how accurately the results of an additive MAUT model are approximated) may benefit from the use of a more concise format. However, an oversummarisation will tend to obscure important information about the riskiness of an alternative. It is difficult to judge *a priori*, but expected values and perhaps standard deviations may fall into this category. Uncertainty formats using “intermediate” levels of detail are therefore expected to perform best.

**Hypothesis 1a:** Uncertainty format has a significant effect on accuracy.

**Hypothesis 1b:** Accuracy will be highest using uncertainty formats that provide “intermediate” levels of detail: three-point (min-median-max) approximations, quantiles, and scenarios.

The difficulty of a decision task is often related to the amount of information processing required to arrive at a choice [213], which in turn is associated with the amount of information provided to the decision maker. The uncertainty formats provide very different amounts of information to decision makers, and it seems likely that those providing more concise information will be judged easier to use. Because of the difficulty involved in reasoning with (e.g. making trade-offs between) variances [103], an exception is expected here: multi-objective decisions involving standard deviations are expected to be relatively difficult despite the conciseness of the information provided.

**Hypothesis 2a:** Uncertainty format has a significant effect on the level of difficulty experienced, with the level of difficulty experienced increasing with the quantity of information provided.

**Hypothesis 2b:** The use of standard deviations will become more difficult relative to the other uncertainty formats as the number of objectives increases.

It is expected that decision makers will tend to focus on information that is particularly salient i.e. that which is directly presented to them [131], and will therefore be more likely to choose alternatives that perform well on the type of information provided by an uncertainty format. Precise hypotheses are listed as Hypothesis 3b, 3c, and 3d. No hypothesis is made involving maxima, because the only uncertainty format which explicitly shows maximum values (the three-point approximations) also shows minimum values at the same time. For most decision makers, the minimum values are expected to be more influential than the maximum values, negating any effect.

**Hypothesis 3a:** Uncertainty format has a significant effect on the type of alternative that is chosen – alternatives with certain characteristics will tend to be selected more frequently using uncertainty formats explicitly showing information on those characteristics.

**Hypothesis 3b:** Alternatives with relatively larger expected values will be more frequently selected if expected values are used.

**Hypothesis 3c:** Alternatives with relatively smaller standard deviations will be more frequently selected if standard deviations are used.

**Hypothesis 3d:** Alternatives with relatively larger minima will be more frequently selected if three-point approximations are used.

## 4.3 Details of choice experiment

### 4.3.1 Recruitment and sample selection

The sample is drawn from postgraduate students (masters and doctoral) at Manchester Business School. Twenty-eight subjects were recruited in September 2008 by email and poster invitations to participate in an experiment and “give some feedback on different ways of presenting uncertainty information in investment decisions”. One group of 13 subjects performed the experiment while carrying out a verbal commentary on their decision-making processes, and another group of 15 simply performed the experiment, without commentary. Subjects were told that the experiment without commentary would take about an hour and would be remunerated at £8, with the experiment with commentary taking about 90 minutes and being remunerated at £12. Subjects were asked to indicate a preference for doing the experiment with or without commentary but were told that their preference may not be accommodated. Most subjects opted for the ‘with commentary’ option, and were allocated to that group randomly.

### 4.3.2 Description of the task

Each subject performs a series of 12 decision tasks, each involving choosing between a set of five investment alternatives. Respondents are informed that, in each problem, they will be selecting an investment for a friend who has been bequeathed £100,000. Depending on the problem, the friend may have one, two or three objectives. These are stated as:

**Objective 1:** Maximise the amount of profit returned after year 1 (to be withdrawn and used to fund a holiday).

**Objective 2:** Maximise the amount of profit returned after year 2 (net of year 1's profit; to be withdrawn and used to fund the purchase of a car). Included in this objective is an additional statement from the friend to the effect that "I'll need at least £4,000 to get any kind of decent car. Anything less than that and all I'll be able to get is a piece of junk".

**Objective 3:** Maximise the total value of the investment after 6 years, net of any withdrawals (to be withdrawn and used to fund the purchase of a house).

If fewer than three objectives are used, they are introduced sequentially in the order indicated above. Respondents are given additional instructions that only one investment may be chosen, and that the weight  $w_j$  that the friend associates with objective  $j$  is given by  $w_1 = 1$ ,  $w_2 = 1$ , and  $w_3 = 2$ . Finally, subjects are told that although the values of the investments at each time period are uncertain, experts have provided some indication of the possible values that each investment might take on, and that different formats have been used to represent this uncertainty.

Six different ways of presenting uncertain outcomes are used, although any one subject will only see two of these: probability distributions, expected values, expected values with standard deviations, three-point approximations/fuzzy numbers, scenarios, and quantiles. Subjects were given a brief (typically about 5 minutes) tutorial on the meaning and use of each of the uncertainty formats relevant to them.

### 4.3.3 Verbal commentary and facilitator involvement

One group of subjects performed the tasks without providing any verbal commentary, and a second group performed the tasks while simultaneously commenting on their decision-making processes. Subjects giving commentary were given a general overview of verbal process-tracing protocols and the reason for verbal protocols in the current experiment i.e. to better understand the processes that people use when making decisions when outcomes are uncertain. They were then asked to provide a running explanation of how they were going about making each decision to the facilitator, with specific emphasis placed on

which pieces of information they were considering at any point in time, and how they were comparing alternatives using this information. Subject were told that any comments on perceived difficulty or simplicity of the problem (or the way in which the problem was formulated), subject fatigue, and interest or boredom were welcome, and that they were also free to make any other comments they wished during the course of the experiment. These commentaries were recorded and later transcribed.

Each subject in the ‘with commentary’ group completed their task in the presence of a ‘facilitator’ (myself). The role of the facilitator was primarily to provide an audience for the subject’s commentary so that the atmosphere is less artificial and approximates that of a conversation, albeit a one-sided one. The facilitator could also prompt the subject to provide more detailed commentary where this was not forthcoming, and direct the conversation to ensure that the essential aspects of the decision process were commented upon. Beyond this, an attempt was made to keep the influence of the facilitator to a minimum.

In both the groups providing and not providing commentary, subjects were advised that at any time they could ask the person administrating the tasks for technical information about the uncertainty formats used. No information about *how* the the decision should be made was disclosed.

#### 4.3.4 Design of investment alternatives

In each decision problem subjects must choose between  $I = 5$  alternatives described on  $J = 1, 2$  or 3 attributes. The performance of investment  $a_i$  on attribute  $c_j$  is denoted by  $Z_{ij}$  and is not known with certainty. The generation of uncertain investment returns uses the same procedure as the simulation study described in Section 3.4.1. Realisations for each  $Z_{ij}$  are drawn from  $L = 3$  normal distributions or ‘futures’, denoted by  $N(\mu_{ij\ell}, \sigma_{ij\ell}^2)$ ,  $\ell \in \{1, \dots, L\}$  where  $\mu$  and  $\sigma^2$  are mean and variance respectively. The  $\mu_{ij\ell}$  are generated on  $U[0, 1]$  and then standardised within each alternative  $a_i$  so that  $\sum_j \mu_{ij\ell} = 1$ , ensuring that the alternatives are Pareto optimal within each future<sup>1</sup>. The  $\sigma_{ij\ell}$  are generated uniformly between 0.01 and 0.1. Thirty values are then generated from each normal distribution  $N(\mu_{ij\ell}, \sigma_{ij\ell}^2)$ . Each  $Z_{ij}$  is therefore represented by 90 realisations. Evaluations on each attribute are then scaled to lie between £0 and £10,000 (for attribute 1 and 2) or between £90,000 and £150,000 (for attribute 3). In certain of the choice problems, one of the generated investment options is replaced by an alternative which has a similar expected value to the investment that it is replacing, but is much less variable on at least one of the

<sup>1</sup>In the case of a single attribute, a different standardisation is carried out (the one described above would just result in values of one for all investments) in which attribute values are standardised over all broad futures rather than within each future.

attributes. The same non-dominance conditions applying to the original set of alternatives are preserved after the inclusion of the less risky alternative.

#### 4.3.5 Uncertainty formats

The manner in which each uncertainty format was presented to subjects is shown in Appendix B.1, which contains a copy of a questionnaire that was presented for one of the two-objective decision problems (six questionnaires are given; one for each uncertainty format). The most salient features of each presentation are given below.

**Probability distributions** Subjects are shown a histogram dividing each attribute domain into ten equal-sized intervals and plotting the relative frequency of each interval. Where multiple objectives are used a grid of histograms is shown, with the investments arranged in columns and objectives in different rows. No numerical information is presented in this format, because it seems certain that any numerical information (even a tabular summary of the information in the histograms, for example) will be so overwhelmingly dense that it will be ignored in favour of the graphical format.

**Expected values** Subjects are shown the expected amount returned by each investment on each objective. Where more than one objective is used, the expected values are shown using in numerical format i.e. in a table, and in a graph in which investments are represented as lines and expected values are plotted on the vertical axis. All objectives are shown on the same table/graph. When three objectives are used, this requires both a left-hand vertical axis (scaled between £0 and £10,000 for the first two objectives) and right-hand vertical axis (scaled between £90,000 and £150,000 for the third objective).

**Standard deviations** Expected values are shown to subjects as before (in both tabular and graphical formats) but in addition standard deviations for each investment are shown in an extra column in the table of expected values.

**Quantiles** This format shows the 5%, 25%, 50%, 75%, and 95% quantiles of each distribution of attribute evaluations. The information is presented as a  $5 \times 5$  matrix of attribute evaluations for each objective, together with an adjoining graph plotting the quantiles along the  $x$ -axis and investment amounts on the  $y$ -axis. Each objective has its own table or graph, with each investment (for a particular objective) represented by a different line in the graph (or row in the table).

**Three-point approximations** Each investment's possible outcomes on an objective are represented by a triple consisting of the minimum, median, and maximum attribute values. These approximations are shown in a tabular as well as graphical format. In



the tabular format, the values are simply listed as minima, medians, and maxima. Then, although it is not strictly necessary to interpret the triples as “fuzzy numbers” (they are simply three-point approximations to the full probability distributions), in the graphical format each graph is shown as a triangular membership function using the  $(x, y)$  co-ordinates: (minimum, 0), (median, 1), and (maximum, 0), with the  $y$ -axis labelled “membership”. Collectively the graphs are displayed in a similar way to the probability distributions (investments in different columns, objectives in different rows).

**Scenarios** Subjects are shown three tables, each one showing the mean performance of the investments on each objective in that future i.e. the  $\mu_{ij\ell}$ ,  $\ell \in \{1, 2, 3\}$ . The evaluations are also presented to subjects as graphs using the same format as the expected value graphs. Prior to the presentation of the results subjects are given a brief description of each scenario as well as an explicit statement that they are considered equally likely to occur. The scenarios are labelled ‘Eastern growth’, ‘Nothing changes’ and ‘Problematic world’ respectively, although their labels and descriptions are theoretically irrelevant to the problem.

#### 4.3.6 Experimental design

The experiment has three treatment effects – uncertainty format (6 levels), number of objectives (3 levels), and the use of commentary (2 levels). The experiment is designed as an incomplete block design in which each subject sees two of the six uncertainty formats and two of the three objective set sizes, and answers three decision problems within each combination of uncertainty format and objective number. Each of the 28 subjects therefore makes 12 investment choices in total, for a total sample size of 336. At an estimated 5 minutes per problem (including the collection of outcome measures), this probably approaches the upper limit of what can be considered a manageable workload for subjects. Originally, it was hoped to have 15 subjects in each of the with- and without-commentary groups, but two subjects scheduled to provide commentary did not arrive for their interviews. Time constraints prevented replacements being found for these subjects. Nevertheless, each uncertainty format is seen by either eight (scenarios), nine (expected values, standard deviations) or ten subjects (all other formats), who each see it in six problems, so that there are between 48 and 60 observations for each uncertainty format. Each combination of uncertainty format and objective set size is seen by between two and four subjects so that there are between 12 and 24 observations for each combination. The 12 problems seen by any one subject are all different problems i.e. subjects do not see the same problem using different uncertainty formats. This means that 18 investment problems are generated in all – six single-objective, two-objective, and three-objective problems – although any one subject will only see 12 of these. The order in which the investment problems are presented

to the subjects is completely randomised.

### 4.3.7 Experimental outcomes

#### Measuring accuracy

As in the simulation study, accuracy is evaluated using “utility loss” [19]. Two utility loss measures are used. The first compares the selected alternative to what would be selected using an idealised additive MAUT model. This is the ‘conventional’ utility loss used in the previous chapter, but it assumes that subjects have access to the full probability distributions used by MAUT. Usually this will not be the case – subjects will see only a summary of the distributions – and it seems somewhat unfair to evaluate subjects based on information that is partly unavailable to them. The second type of utility loss thus compares the selected alternative to what would be selected using an idealised application of whatever simplified value function model is appropriate for the uncertainty format presented in a problem (e.g. an expected value model, a quantile model). The same simplified models are used as in the simulation reported in Chapter 3; see Appendix B.2 for details.

The use of both types of utility loss requires that one specifies the attribute weights and marginal utility functions to be used as inputs to the idealised (MAUT or simplified) models. Idealised weights are those specified by the problem description ( $w_1 = w_2 = 1$ ,  $w_3 = 2$ ), but the precise shape of the idealised marginal utility functions cannot be known with certainty. A number of functional forms that appear plausible based on the problem description and an assessment of subjects’ statements of their decision-making strategies have therefore been used. These indicate two dominant features: a strong tendency towards risk aversion i.e. concave utility functions, for attributes 1 and 3; and a very strong preference threshold around £4,000 for attribute 2. A full list of the utility function parameters used for 12 plausible preference structures is given in Appendix B.3. Each of the functional forms results in an estimate of utility loss. These 12 utility losses are averaged to obtain a final estimate of utility loss i.e.  $UL = \sum_{j=1}^{12} (U_{i^*}^{(j)} - U_{i^{sel}}^{(j)}) / (U_{i^*}^{(j)} - U_{i_*}^{(j)}) / 12$  where  $U_{i^*}^{(j)}$  is the utility (obtained using idealised model  $j$ ) of the best alternative in idealised model  $j$ ’s rank order,  $U_{i_*}^{(j)}$  is the utility of the worst alternative in idealised model  $j$ ’s rank order, and  $U_{i^{sel}}^{(j)}$  is the utility (again according to idealised model  $j$ ) of the alternative selected by the subject. The sensitivity of the accuracy results to preference structures is shown in Appendix B.4, which presents accuracy results separately for each of the sets of marginal utility functions. The accuracy results of the expected value model can vary considerably depending on the form of the underlying utility functions – they are excellent when utility functions are linear but deteriorate relative to the other formats as preferences become more non-linear. The main conclusions do not vary greatly depending

on which of the utility functions in Appendix B.3 are used.

### **Measuring perceived difficulty and confidence in answer**

Following each investment decision subjects were asked to indicate how difficult they had found it to arrive at a decision. Their answers were given on a scale from 1 (extremely easy) to 7 (extremely difficult). Subjects were also asked to indicate how confident they were of having chosen the best of all the investments, by completing a statement “Given 100 problems similar to this one, I would expect to choose the best investment in about  $x$  of them”. Subjects were specifically instructed that they should consider both their confidence in their decision-making and the fact that they may not have had access to all information. After answering all 12 decision problems, subjects were given a final survey in which they indicated which of the two uncertainty formats they felt helped them to make better decisions, and which they found easier to use. This set of questions was asked separately for both numbers of objectives that the subject had seen. Subjects were given an option of indicating that no difference was perceived between the uncertainty formats.

### **Measuring the characteristics of the chosen alternative**

For each choice task, different alternatives possess different characteristics e.g. a large expected value, a small standard deviation. The selection of an alternative can be interpreted as providing an indication of the desirability of those characteristics. The approach used here is to describe the performance of an alternative on each attribute in terms of its rank among the set of alternatives according to four measures: expected values, standard deviations, minima, and maxima. Descending rank orders are used for expected values, minima, and maxima (where larger values are preferred); an ascending rank order is used for standard deviations (where less is preferred). A rank of one thus indicates the most desirable alternative on any characteristic. For problems involving multiple objectives, alternatives’ characteristics can be described for each objective and final rank measures are calculated by averaging the ranks obtained for each attribute.

#### **4.3.8 Model building**

For the outcomes described above, one might expect responses to be correlated within subjects – for some subjects to exhibit generally higher quality decisions, or to tend to choose alternatives of a particular kind, etc. To account for the possibility of correlated observations, generalised estimating equations (GEE’s, e.g. [173, 302]) has been used to fit models to the data. These assume that the marginal distribution of a dependent variable follows a generalised linear model (GLM’s; e.g. [190])  $g(E[Y_{ij}]) = \mathbf{x}_{ij}\beta$ , where  $Y_{ij}$  denotes the value of outcome  $j$  for subject  $i$ ,  $\mathbf{x}_{ij}$  is a vector of explanatory variables, and  $\beta$  is a vector of parameters. The covariance structure for the vector of observations within a par-

ticular subject  $i$ , denoted  $\mathbf{Y}_i$ , is given by  $\phi \mathbf{A}_i^{1/2} \mathbf{R}(\alpha) \mathbf{A}_i^{1/2}$ , where  $\mathbf{A}_i$  is a diagonal matrix of variance functions (in the quasi-likelihood approach employed by GLM's and GEE's, these are known functions of the mean response i.e.  $h(E[Y_{ij}])$ ), and  $\mathbf{R}(\alpha)$  is the so-called “working” correlation matrix of  $\mathbf{Y}_i$ . The vector of parameters  $\alpha$  specifies the nature of the correlation structure – for example, the standard GLM assumes that repeated observations for a given subject are independent of one another – and the matrix is referred to as a “working” one because the parameter estimates and their estimated variances remain consistent even if the correlation structure is misspecified, provided that the model for the mean is correct [302]. Because the order of task presentation is completely randomised, a simple “exchangeable” correlation structure is used which assumes a constant correlation between any pair of observations within the same subject i.e.  $\text{corr}[Y_{ij}, Y_{ik}] = \alpha$  for all  $j \neq k$ .

For responses involving difficulty, confidence, and the ranks of the selected alternative according to certain characteristics, Gaussian probability functions with identity links are used. The accuracy outcomes are fractional response variables for which a logit link function and binomial probability function can be used whenever a quasi-likelihood approach is adopted [210], as is the case with GEE's. Two-factor interactions were considered and have been included where significant at the 10% level. Results for main effects are shown whether they are significant or not. The effect of uncertainty formats are shown using probability distributions as a reference category. On a final note regarding outliers and data cleaning, one subject, when presented with a single-objective problem using expected values, selected the investment with the *lowest* expected value in two of the three problems (and the investment with the second-lowest expected value in the third problem). The single-objective expected value problem is seemingly trivial, since each alternative is represented by only a single number (to be maximised), and thus the choices are highly suspect. Since the study is partly concerned with errors using decision aids, one is loathed to remove observations where errors have clearly been made. However, it seems that here the error cannot be expected to generalise and is simply the result of carelessness. I argue that this carelessness would be far less likely to happen in real-world decision making, and therefore delete these three observations from the analysis. It is worth noting, since one might suspect the entire set of 12 choices made by that subject, that further deletion of observations had no effect on any of the conclusions drawn.

## 4.4 Results

### 4.4.1 Factors affecting accuracy

Mean utility losses are 0.20 if accuracy is computed relative to a MAUT model and 0.15 if it is computed relative to an idealised simplified model (standard errors of both means

are less than 0.01). Median responses are 0.09 and 0.02 respectively. Mean MAUT utility losses are 0.26 for expected values, 0.23 for standard deviations, 0.25 for probability distributions, 0.12 for quantiles, 0.15 for three-point approximations, and 0.17 for scenarios (all standard errors are 0.03). These figures indicate that subjects generally performed far better than random guessing, which returns a MAUT utility loss of 0.5. Figure 4.1 shows parameter estimates for variables potentially influencing accuracy, with estimates for each uncertainty format ('EV' = expected values; 'SD' = standard deviations; '3pt' = three-point approximations; 'Quan' = quantiles; 'Scen' = scenarios) shown in Figure 4.1(a) and estimates for other effects ('Rep' = number of previously seen problems of this type; 'Mcr' = more than one objective; 'Comm' = commentary given) shown in Figure 4.1(b). In Figure 4.1(a) parameter estimates should be interpreted relative to the appropriate reference category – probability distributions. Effects on utility loss calculated using a MAUT model are shown on the  $x$ -axis, and effects on utility loss calculated using an appropriate simplified model are shown on the  $y$ -axes. Labels are plotted at the estimated parameter values, with 90% confidence intervals indicated by dashed ellipses. Statistical significance (at the 10% level) can be deduced from the ellipses.

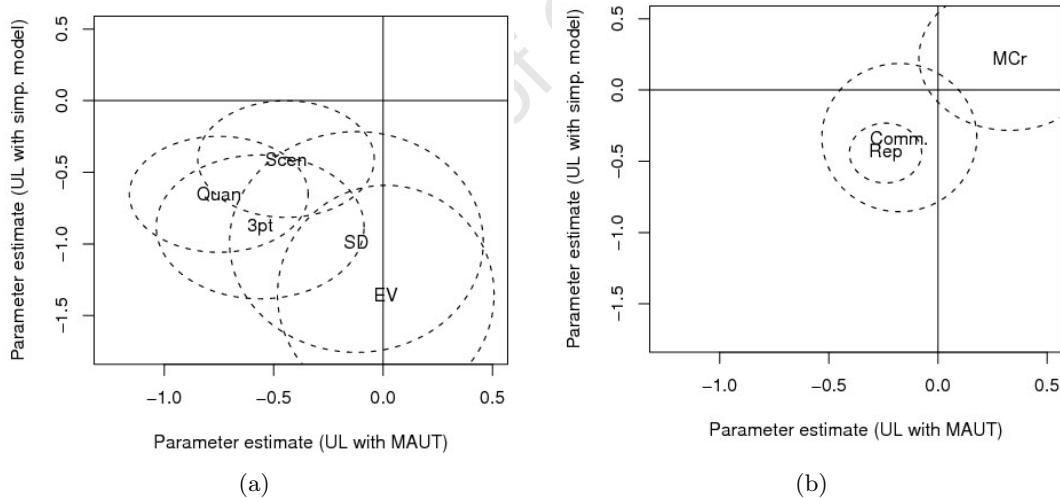


Figure 4.1: Parameter estimates for models of decision accuracy. The effects of uncertainty formats are shown in (a), other effects are shown in (b).

Uncertainty formats that lead to more accurate approximations of MAUT than probability distributions are those lying to the left of  $x = 0$ , and those that lead to good decisions according to an idealised application of their own simplified models are those lying below the line  $y = 0$ . Results indicate that quantiles, three-point approximations, and scenarios all give significantly lower MAUT utility losses than probability distributions. Standard deviations and expected values do not differ significantly from probability distributions in their ability to approximate MAUT. These results support both Hypothesis 1a and

Hypothesis 1b.

The results obtained relative to the idealised application of simplified models – those shown on the  $y$ -axis – indicate that subjects are able to use standard deviations and expected values to choose alternatives that are relatively good according to models considering only standard deviations and expected values. The poor approximations of MAUT obtained using these two uncertainty formats are therefore more due to limitations in the amount of information that they can convey than to subjects using the information in a sub-optimal way (the opposite is true of probability distributions). In contrast the location of the ellipses for scenarios, quantiles, and three-point approximations in the lower-left quadrant of Figure 4.1(a) indicates that subjects are able to make relatively good use of the information conveyed by those formats and that those models are also capable of closely approximating MAUT results.

Figure 4.1(b) shows the remaining effects. Decision accuracy improves as subjects see more problems of a similar type, which would seem to indicate some kind of ‘learning’ or familiarisation process. Accuracy also shows a very marginal tendency to deteriorate as the number of objectives increases or if commentary is not provided – although it must be stressed that these effects are not significant at the 10% level. Accuracy results show essentially no within-subject correlations (within-subject correlations are 0.04 (s.e. 0.03) and 0.05 (s.e. 0.03) for accuracy results using MAUT and simplified models respectively).

#### 4.4.2 Factors affecting assessments of difficulty

The mean difficulty rating given by subjects is 3.21 (s.e. 0.07), indicating that most subjects found the task moderately easy. Mean difficulty ratings are 2.9 for expected values, 3.8 for standard deviations, 3.5 for probability distributions, 3.5 for quantiles, 2.8 for three-point approximations, and 2.7 for scenarios (standard errors are between 0.16 and 0.21). Confidence in the quality of answer given is correspondingly high, with an average evaluation of 76 (s.e. 0.8). Some 10% of all confidence ratings indicated 100% confidence, which should only be possible under perfect information and a perfect decision-making strategy (this was emphasised to subjects before the task). This suggests that either some subjects were vastly overconfident in their abilities or (probably more likely) that the 0-100 scale was interpreted in a subjective rating-scale sense rather than the intended probabilistic one. Figure 4.2 shows parameter estimates for variables potentially influencing perceptions of difficulty and confidence, with the same format as Figure 4.1 except that the  $x$ - and  $y$ -axes shows effects on perceived difficulty and confidence respectively. Uncertainty formats that are evaluated as easier to use than probability distributions are those lying to the left of  $x = 0$ , and those that lead to more confidence in decisions are those lying above the line  $y = 0$ .

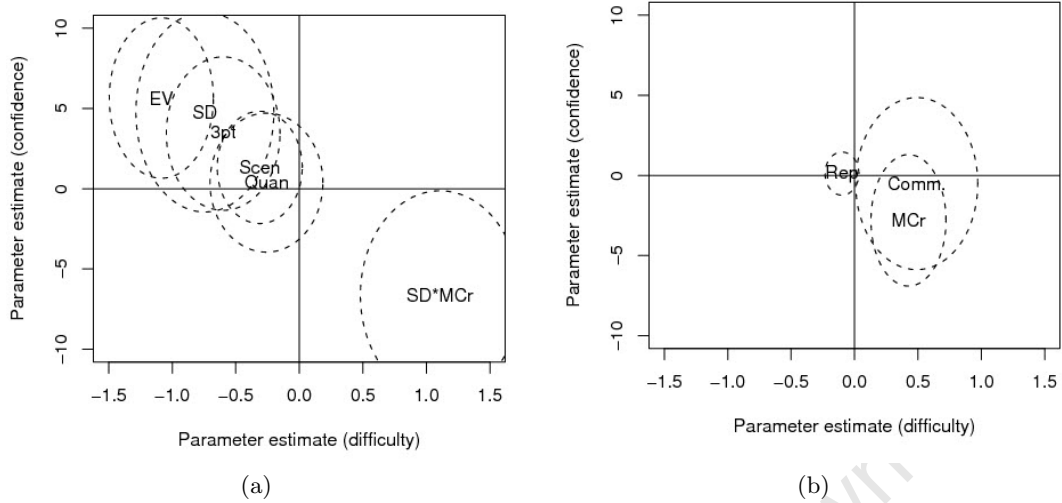


Figure 4.2: Parameter estimates for models of perceived difficulty and confidence. The effects of uncertainty formats are shown in (a), other effects are shown in (b).

Figure 4.2(a) shows a strong inverse relationship between difficulty and confidence, supporting the suggestion that when evaluating their confidence subjects did not take into account the amount of information that is lost when using a simplified model but rather based both responses on difficulty. Subjects found expected values, standard deviations (in single-objective problems), and three-point approximations significantly easier to use than probability distributions. Coefficient signs for the quantile and scenario formats are also negative but are not large enough to be significant, even at the 10% level. The strong positive interaction between the use of standard deviations and the presence of multiple objectives indicates that the use of this format becomes relatively more difficult when multiple objectives are present, possibly because subjects find it difficult to make trade-offs between the more abstract variance information. Both Hypothesis 2a and 2b are thus supported by the data. Confidence evaluations, although less sensitive to uncertainty format, show essentially the same pattern. Only the expected value effect and the interaction between standard deviations and multiple objectives are large enough to be significant at the 10% level.

Subjects giving commentary rated the task as significantly more difficult than those who did not give commentary. Subjects also indicated that problems became significantly more difficult when multiple objectives were present and trade-offs between attributes had to be considered. A further model was fitted using indicator variables to separate the effect of two-objective and three-objective problems (relative to the reference single-objective case). Both effects were significant at the 5% level and were of similar magnitude (0.41 for the two-objective indicator, 0.52 for three objectives), indicating that the increase in

perceived difficulty in going from one to two objectives was considered to be substantially more than that experienced in going from two to three objectives. The effect of seeing one or two problems of a similar type did not exercise a significant influence on difficulty. In contrast to the decision accuracy results, evaluations of difficulty show substantial positive within-subject correlation (0.31 (s.e. 0.05) for difficulty; 0.31 (s.e. 0.07) for confidence). This indicates some general subject-specific differences, although because of the use of a constructed difficulty scale, it is not possible to say whether some subjects genuinely felt the tasks were more difficult than others or were interpreting the scale in a different way.

It is useful to combine the difficulty and utility loss results to broadly sketch the properties of different uncertainty formats for multi-objective decision making. Figure 4.3(a) shows the predicted MAUT utility loss and the predicted difficulty rating for decision problems with multiple objectives. These evaluations can be compared with the overall ratings of quality and difficulty gathered at the very end of the set of choices, shown in Figure 4.3(b). This figure shows, for each uncertainty format involving multiple objectives, the proportion of all pairwise comparisons involving an uncertainty format in which subjects perceived it as leading to worse decisions (as an inverse measure of accuracy, comparable to utility loss) or being harder to use (as a measure of difficulty). The dashed ellipses indicate 90% confidence intervals around the predicted values and proportions in Figure 4.3(a) and (b) respectively.

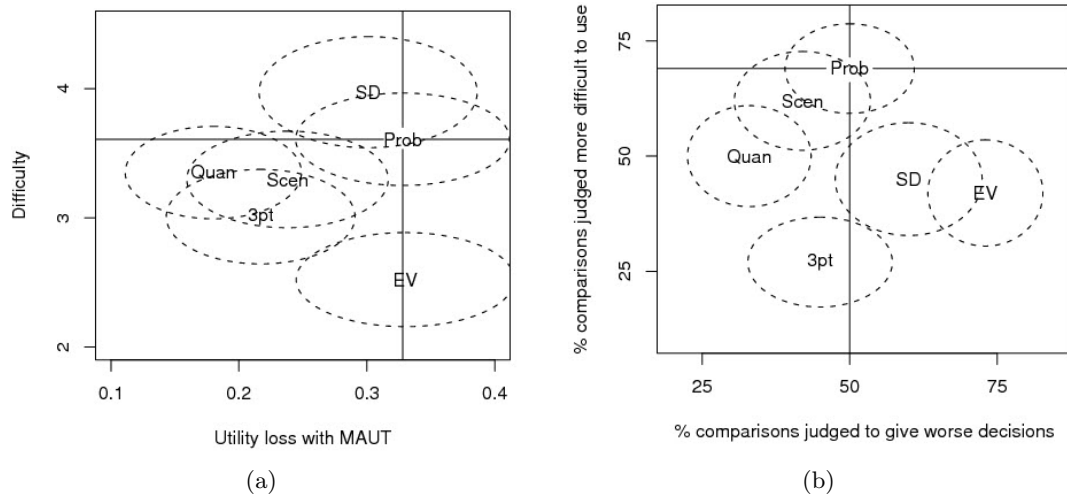


Figure 4.3: Modes of uncertainty format evaluated according to their accuracy and difficulty in multi-objective problems. In (a) accuracy and difficulty are predicted values obtained from the fitted models; in (b) they are obtained by holistic judgements made by subjects after the set of tasks was completed.

Figure 4.3(a) clearly shows the relative attractiveness of quantiles, three-point approxima-



tions, and to a lesser extent scenarios. Quantiles offer slightly greater accuracy than three-point approximations, but are viewed as more difficult to use. Probability distributions and standard deviations are essentially dominated by scenario assessments, three-point approximations, and quantiles. The positioning of expected values as an easy-to-use but relatively poor-quality decision aid is also clear. The overall evaluations shown in Figure 4.3(b) are largely consistent with the predictive results shown in Figure 4.3(a). Two of the main results – the accuracy of decisions using quantiles to present uncertainty, and that three-point approximations offer additional ease-of-use with some trade-off in accuracy – are clearly visible in Figure 4.3(b). The only noteworthy differences are that probability distributions tend to be rated as relatively more accurate, and standard deviations and three-point approximations relatively easier to use, in the holistic comparisons shown in Figure 4.3(b).

#### 4.4.3 Factors affecting the type of alternative chosen

The average rank of the selected alternative in rank orders of standard deviations and minima – characteristics indicative of lower risk – are 2.26 (s.e. 0.05) and 2.14 (s.e. 0.05) respectively. This is somewhat lower than the average rank in rank orders of expected values (2.64, s.e. 0.05) and maxima (3.22, s.e. 0.07), indicating that subjects tended to choose less risky alternatives more often. Figure 4.4 shows parameter estimates for variables potentially influencing the type of alternative chosen. In Figure 4.4(a) the  $x$ - and  $y$ -axes show the relative effect of uncertainty format on the rank of the selected alternative according to expected values and maxima respectively. In Figure 4.4(b) the  $x$ - and  $y$ -axes show the effects on the rank according to standard deviations and minima respectively. Uncertainty formats that tend to favour the selection of alternatives with a particular characteristic will be located to the left of  $x = 0$  (in the case of expected values and standard deviations) and below  $y = 0$  (for maxima and minima).

There are three main results in Figure 4.4(a) and (b), which apply only (or to a far greater extent) in single-objective problems. Firstly, subjects shown expected values are more likely to choose alternatives with large expected values or maxima, and less likely to choose alternatives with small standard deviations or large minima. Secondly, subjects shown standard deviations are less likely to choose alternatives with large maxima. Thirdly, subjects are more likely to choose alternatives with relatively large minima if three-point approximations are used. The latter two effects disappear when there are multiple objectives, and the first effect is weaker – subjects are no longer more likely to select alternatives with large expected values, are only slightly more likely to select alternatives with large maxima; but they remain less likely to choose alternatives with small standard deviations and large minima. It therefore appears as if the presence of trade-offs

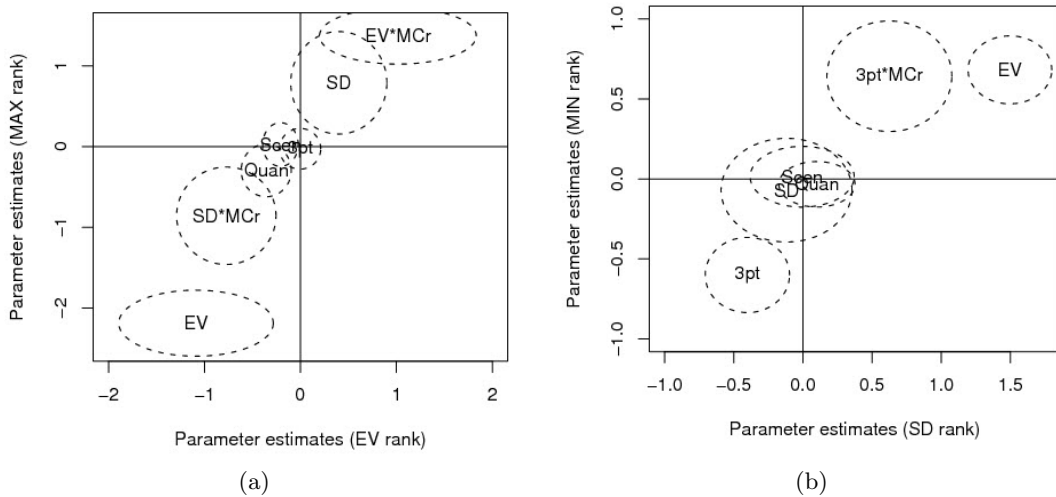


Figure 4.4: Parameter estimates for the rank of the chosen alternative according to specific characteristics: (a) expected values and maxima, and (b) standard deviations and minima.

between objectives plays a role in drawing attention away from the pure “risk” aspect of the problem.

With regard to the research hypotheses, at best mixed support is found. Alternatives with larger expected values are more likely to be selected if expected values only are shown (Hypothesis 3b), but only in the trivial single-objective case where this is the only information given. There is no tendency for subjects to select alternatives with small standard deviations relatively more often if standard deviations are shown (contrary to Hypothesis 3c). Alternatives with larger minima are more likely to be selected using three-point approximations (Hypothesis 3d), but again only in the single-objective case (although this is no longer trivial).

The remaining effects are much smaller in magnitude, and are not shown in detail here. Subjects are less likely to choose alternatives with large expected values in multi-objective problems ( $\beta = 0.28$ ,  $p = 0.03$ ), but are more likely to do so after seeing a problem type several times ( $\beta = -0.13$ ,  $p = 0.02$ ). Subjects are more likely to choose alternatives with large minima ( $\beta = 0.54$ ,  $p < 0.01$ ) in multi-objective problems. There is essentially no within-subject correlation in any of the four outcomes (within-subject correlations are between -0.02 (s.e. 0.02) and 0.04 (s.e. 0.04)).

#### 4.4.4 Analysis of verbal commentaries

In this section, a qualitative summary of the subjects’ verbal commentaries is given. An attempt is made to identify those features of subjects’ decision making processes that arose regularly and seem to be of particular interest.

### **Salience of problem description**

One aspect of the problem description that was frequently mentioned by subjects was the indication that more than £4,000 was required for the car to avoid buying a “piece of junk”. All of the 13 subjects made some reference to satisfying this goal at some stage in their commentaries; 10 subjects referred to it in more than a third of the (two- or three-objective) problems that they saw, and 6 subjects referred to it in more than half the problems that they saw. Subjects made use of this information in two quite different ways. Most frequently, the fact that an alternative did not guarantee (or almost guarantee) an amount of £4,000 for the car was used as a basis for excluding it. This was often done in an initial ‘screening’ phase of evaluation [e.g. 196, 21]. Otherwise the probability of obtaining £4,000 was used as a measure of the attractiveness of an alternative i.e. as an explicit risk attribute. When scenarios were used, the number of scenarios in which the amount received for a car exceeded £4,000 was used in a similar fashion.

A second feature that appears regularly in subjects’ commentaries and also involves the salience of problem descriptions is a propensity to begin the decision making process by considering performance on the most important objective, namely the house. This behaviour can only be observed in three-objective problems, but in those cases it is fairly widespread. Ten of the 13 subjects saw tasks involving the house objective, and of those ten only two never considered the house objective first. Four of the remaining eight subjects considered the house objective first in more than 50% of the problems they saw, although only one subject did this in every problem. However, recognising that one objective is most important does not always imply that it should be considered first – this largely depends on what choice strategy is being used. Most compensatory approaches – like maximising expected value or utility – do not require that objectives be considered in any particular order. However, satisficing and other lexicographic strategies do imply that objectives are considered in the order of their importance to the decision maker. On this point, it is worth noting that the three subjects who considered the house objective first also made the most frequent use of a satisficing-like heuristic (discussed in the next subsection).

### **Decision-making strategies**

Subjects often worked by sequentially eliminating alternatives from consideration. This elimination serves two purposes. First, it can be used in the early part of the decision process to remove alternatives which are considered to be clearly unsuitable. This allows the decision maker to then focus on a smaller number of alternatives in more detail. This behaviour is widespread even though the number of alternatives is already quite small to begin with. Second, some subjects continue with sequential elimination of alternatives until only one remains. Naturally, it is not always clear when one process (initial removal

of unsuitable alternatives) ends and the other (a more subtle exclusion of alternatives) begins. Alternatives are most often eliminated because they (a) do not offer £4,000 for the car with sufficient certainty; (b) are too risky in the sense that they offer an unacceptably high chance of getting a zero or near-zero amount on the holiday or car objectives, or (c) are too risky in the sense that they offer an unacceptably large range of outcomes on one or more objectives. Interestingly, although subjects often excluded alternatives because of a possible zero-amount on the holiday or car objective, no subject explicitly excluded an alternative because of a possibility of getting £90,000 for the house objective – even though this is the minimum possible. This suggests that the value zero formed a particularly salient reference point for subjects' judgements i.e. the situation in which *nothing* is obtained and no purchase (of a holiday or of a car) is possible at all.

Although the elimination of alternatives is by some way the most frequently observed element of decision making strategy emerging from the commentaries, other features can also be discerned. Where attractive features of an alternative's performance have been mentioned, these most often relate to (a) a high minimum value, (b) a small range between worst and best outcomes, and (c) a good "balance", particularly between the amount received for the car and holiday. The first two of these features are fairly obvious in light of the features perceived as particularly negative. The search for a "well-balanced" alternative was often justified using the equal importance of the car and holiday objective, but even in the absence of this justification there was a strong tendency for subjects to view alternatives doing excellently on one objective (even the house objective) at the expense of the other objectives to be relatively poor.

There is fairly frequent use of a strategy bearing a strong resemblance to a satisficing-type heuristic. The strategy involves the identification of a single alternative that does best on some objective, generally maximising the value for a house. Following this, the selected alternative's performance on other attributes is assessed. If the performance is satisfactory, the alternative is selected; if not, the alternative that is next-best on the original attribute is selected and the process continues. The checking of performance on other attributes is usually done in a fairly holistic way, and the subjects using this strategy differ quite strongly with respect to how good the performance must be in order to be accepted. One subject performs only a cursory inspection, another ends up always rejecting the first selected alternative, from which one can assume that he is being fairly demanding. One particularly clear example of this strategy is provided by a subject who uses the 50% quantile of performance on the house objective to select an alternative. Performances on the 5% and 95% quantile of the house objective are then checked. Following this, performance on other objectives is evaluated in a more holistic way – no details are provided in the comments beyond "good", "bad", "unacceptable", and so on.

The “satisficing” heuristic described above is observed in three of the 13 subjects at some stage of their decision making. Other heuristics that are also found at some stage include: the “max-min” heuristic in which the alternative with the largest minimum value (on a particular attribute, or across attributes in the multi-attribute case) is selected, the “take the best” heuristic in which the alternative doing best on the most important attribute is chosen; the “elimination by aspects” heuristic in which alternatives are removed if they fail to achieve a certain performance level (until only one remains), and the “majority of confirming decisions” heuristic in which the alternative that does best on the greatest number of attribute is selected. In only one case does a subject explicitly attempt to calculate the expected value to be obtained from the various alternatives, and to choose the alternative with the highest expected value.

### **Use of uncertainty information**

Subjects seeing three-points approximations often referred to minimum evaluations, using these especially in the initial phases of the decision process to exclude alternatives which had small minima on one or more objectives. This was perhaps the clearest influence of uncertainty format on subjects’ decision-making processes. Subjects using probability distributions were observed to make reasonably frequent use of the mode of the distribution. This was done without explicitly labelling it as such, but by making comments regarding the most likely outcome or outcomes. Although the median is of course something quite different, subjects using three-point approximations made less frequent comments about this “middle” quantity – even though they only had three measures of attribute performance to begin with. Variance information was used in two quite different ways. Two of the four subjects seeing standard deviations used them to make very holistic descriptions of an alternative’s desirability. In one case, this was accompanied by explicit comments that it was difficult to work with standard deviations, so one interpretation of these holistic judgements is that they are best attempts in the absence of a clear idea of how to use standard deviations. The other two subjects constructed ranges by adding and subtracting some multiple of the standard deviation from the expected value. These ranges were then used as effective minimum-maximum ranges.

### **General comments**

Most subjects giving commentary displayed a surprising level of enthusiasm and commitment to the tasks. No real boredom or fatigue seemed to creep into their analyses, and no comments to this effect were made. All subjects requested that the results of the study be sent to them. Several subjects indicated that they thought that the context of the decision problem i.e. investments, had a significant influence on their decisions, and that their responses may have been quite different in contexts not specifically involving investments.

When pressed however, subjects were not really able to elaborate on how or why their decisions might have been changed by a different context. A related problem was that subjects often attempted to adjust their friend’s weighting of the three objectives in line with their own preferences – almost always by trying to impose that the house should be the overwhelmingly more important objective (because it “appreciates in value over time”, or is “a fixed asset”, and so on). Every attempt was made to focus subjects’ attention on the stated preferences rather than their own, but probably not with total success.

## 4.5 Discussion

In Chapter 3 a simulation experiment was used to evaluate the potential ability of a number of simplified decision models to replicate the results obtained using MAUT. The main messages emerging from those results were that an idealised application of a simplified model is on average capable of producing results close to an idealised application of MAUT, and that a quantile model gives consistently better approximations than any other simplified model. One concern with the simulation results is whether they carry over into real-world prescriptive decision making. Another is that they are incapable of evaluating critical practical issues for decision aid – ease of use, for example. In this chapter, by considering actual choices made by unfacilitated subjects in a controlled experimental setting, some progress has been made toward addressing these concerns.

### 4.5.1 Reconciling the simulation and choice experiments

A simplified model’s approximation of MAUT would be worse in practice than when simulated if the uncertainty format used by the model biases the selection process in favour of weaker (according to MAUT) alternatives. Biases may lead to some uncertainty formats being preferable to others, or to the creation of tools for removing as much of the bias as possible. Evaluating the biases in prescriptive decision making is difficult or impossible. The time and effort involved in sourcing and assessing enough decision problems for any finding to be statistically reliable is prohibitively large, but even a single within-subject comparison requires that the same decision maker independently construct facilitated MAUT and simplified models, which in itself seems difficult to achieve.

In the absence of prescriptive results, some progress can be made by examining unfacilitated decisions and identifying any biases that may exist when using different uncertainty formats. There is good reason to suspect that descriptive biases may also arise in prescriptive decision making unless they are specifically addressed by the prescriptive analysis. This sentiment has a long history in decision analysis in the context of probability elicitation (e.g. [102]). But because descriptive biases may be reversed with the facilitation, additional effort and motivation that would usually accompany a prescriptive process, any

prescriptive conclusions must be drawn tentatively.

There is nothing in the results of this chapter to suggest that radical changes to the main conclusions drawn by the simulation study are required. Results continue to suggest that the amount of information lost when using a simplified model does not prevent results from approximating MAUT fairly closely. In the choice experiment average utility losses ranged between 0.12 for quantiles and 0.26 for expected values, well below the 0.5 expected from a random alternative selection. These utility losses were achieved without facilitation and using artificial decision problems of no real consequence to the decision makers. Better utility losses might be expected in real-world prescriptive decision making. The utility loss results are roughly comparable in magnitude to those obtained using simulation with large (30%) assessment errors, where the utility losses of the 5-quantile and expected value models were 0.11 and 0.19 respectively (although the scenario model with 100% coverage performed relatively better in the simulation – utility losses were 0.09 and 0.16 in the simulation and choice experiments respectively).

The main result is that uncertainty format can exert a significant influence on the unfacilitated choice of an alternative. For the tasks performed in this study and for the set of participants used, the use of probability distributions appeared to overload subjects, leading to relatively poor choices. This finding cannot be directly compared to any of the simulation results, which only consider the idealised use of probability distributions in a MAUT model. From a prescriptive perspective the results highlight the need for considerable facilitation and model building when probabilities are used to represent uncertainty, because in the absence of facilitation decision makers use probability information poorly.

Of the other uncertainty formats, selected alternatives most closely approximated those preferred under MAUT if quantiles, scenarios, or three-point approximations were used. All of these formats provide an intermediate level of detail when summarising probability distributions. Expected values and standard deviations returned significantly lower accuracies. By comparing the selected alternatives with what would have been chosen using an idealised application of simplified models using expected values and standard deviations, subjects were shown to be able to use this information effectively, choosing alternatives that were good according to the simplified models. The poor approximations of MAUT obtained using expected values and standard deviations are therefore predominantly due to limitations in the amount of information that they can convey.

The experimental results also support the conclusion that a quantile model gives consistently better approximations of MAUT than other models, which was drawn from the simulation results. Although fuzzy decision models performed worse relative to other uncertainty formats in the simulation experiment, in the absence of facilitation the three-

point approximations may have also been used as quantiles. Simulation results also indicated the scenario models could give excellent approximations if all futures are captured in the scenarios, as is the case here. The relatively poorer performance of standard deviations and to a lesser extent expected values is also in basic agreement with the simulation results.

#### 4.5.2 Practical issues

Because the choices in this chapter are real rather than simulated, the effects of uncertainty format on more practical aspects of decision making can also be investigated – particularly ease of use and the mental processes by which decisions are reached. Table 4.1 summarises the main results and provides a classification of the six uncertainty formats in terms of their effects on the main experimental outcomes.

Subjects found the detailed probability distributions relatively difficult to use, and more concise formats – particularly expected values, standard deviations (in single-objective problems), and three-point approximations – relatively easy to use. Standard deviations were viewed as relatively more difficult to use in problems with multiple objectives, suggesting that trade-offs involving standard deviations are difficult to think clearly about. For decision making with multiple objectives, the use of probability distributions and standard deviations is dominated by – in the sense of leading to poorer and more difficult decisions than – the use of quantiles, three-point approximations, and scenarios. Expected values occupy a unique position in that they are easy to use but lead to relatively poor decisions. From a prescriptive perspective, the difficulty in using probability distributions – which appears to be due to the amount of information given and a lack of ideas about how to structure it – would probably be substantially reduced with facilitation. In the case of standard deviations, relatively little information is shown and difficulty is experienced not because of any overload but because of the difficulty of understanding and reasoning with standard deviations. This difficulty, which is conceptual rather than due to a lack of structure, would probably be reduced by facilitation but not to the same extent as probabilities. The difficulty of using probability distributions may also be justified by the improved decisions that surely arise when these are facilitated, but as the simulation results have shown even the idealised application of a variance model returns relatively poor selections. The experimental and simulation results suggest that the standard deviation is a relatively poor way of representing uncertainty in prescriptive decision problems.

Uncertainty formats convey information about the full probability distributions using different quantities. Drawing attention to certain quantities rather than others can not only affect the choice of final alternative and the ease of use and understanding, but also the process by which a choice is made. In this experiment decision makers shown minimum



evaluations tended to focus strongly on these, selecting alternatives with significantly larger minima in single-objective problems and eliminating alternatives that had low minima on one or more attributes. This behaviour occurred even though no information was given about how likely obtaining this minimum was. Attention was especially strongly drawn to the minimum when this value was zero. Decision makers tended to be less extremely risk averse if 5% and 95% quantiles were used instead of the extremes, partly because some probability information is provided. However, given that the main purpose of a prescriptive process is to provide a structure to replace many of these descriptive heuristics, one might expect this increased risk aversion induced by the minima to be removed or much reduced by the construction of utility functions in prescriptive applications.

University of Cape Town

Effect on	Probabilities	Expected values	Std. devs	3-point/Fuzzy	Quantiles	Scenarios
Decision quality	Worse than three-point approx, quantiles and scenarios	Relatively poor. Subjects to apply model well but valuable information is lost by ignoring uncertainty	Same as probability distributions.	Second-highest, after quantiles. Significantly better than probability distributions.	Best choices. Significantly better than probability distributions.	Slightly but not significantly better than probability distributions.
Perceived difficulty	Relatively difficult. Significantly harder than EV and three-point approx.	Rated easiest to use.	Easier than probability distributions with one objective but more difficult if multiple objectives.	Significantly easier than probability distributions.	Slightly but not significantly easier than probability distributions	Slightly but not significantly easier than probability distributions
Type of investment chosen	Conservative alternatives (large min or small $\sigma$ ).	Risky alternatives (large maxima and EVs, smaller min-ima and larger var)	Slightly more conservative alternatives than probability distributions in one-objective problems (smaller EV and maxima)	Highly conservative alternatives in one-objective problems (tends to a max-imin strategy)	Slightly riskier alternatives than probability distributions (larger EV and maxima)	Very similar to probability distributions.

Table 4.1: A classification of the experimental uncertainty formats according to their effect on naïve decision making

## Chapter 5

# Applications of decision support with simplified representations of uncertainty

### 5.1 Introduction

The two previous chapters have evaluated simplified uncertainty formats in the artificially induced conditions of simulation and choice experiments. The controlled nature of these experiments allow different uncertainty formats to be directly compared and provide a number of insights, but are limited in the practical issues that they can address. In this chapter three real-world applications involving decision analysis with simplified uncertainty formats are reported. The aim of the chapter is (a) to establish whether a simplified decision model can be employed in a practical decision problem; (b) to identify any difficulties experienced during the application, and (c) to establish whether the additional decision support provided by incorporating assessments of uncertainty is perceived by decision makers to be useful. Of course, in addition to these “research” objectives, in each application decision makers will have their own aims for the analysis. In this respect I follow the distinction drawn in the action-research literature (e.g. [40]) between the “action” aims of the decision maker and my own “research” aims.

In the first problem, a group of energy researchers must decide which of a number of electricity saving options should be advocated to the public as part of a media campaign aiming to reduce household electricity consumption. In the second problem, the management of a casino wish to evaluate their portfolio of slot machines and assess which of these machines are dominated or nearly-dominated. In the third problem, a group of market researchers must evaluate how one “client” alternative performs in relation to its competitors on the basis of multi-attribute evaluations collected in a survey. Attribute evaluations in all of the problems turn out to be uncertain, and in all cases the use of probability distributions was not considered appropriate. This provides an opportunity to

use and evaluate some of the simplified uncertainty formats described in previous chapters.

All three decision problems are somewhat unusual for decision analysis. The first two involve providing support for groups tasked with selecting a single shortlist of alternatives on behalf of individuals belonging to a market or ‘population’ for which the group is responsible – each of the individuals making up the market will then make their own un-facilitated choice from the shortlist. In the third application, the group receiving support is tasked with providing information about a particular alternative for which the group is responsible. In all three applications, the group for whom decision support is provided is not the decision maker (in the sense of the people whose preferences it is wished to model), but may be thought of as effectively acting as an ‘expert’ providing what knowledge it can about the decision problem and the decision makers. Furthermore, very little is known about the preferences of the individuals making up the market, and no direct contact is made with them to assess this information.

Although not by construction, the lack of decision maker involvement in each of these problems made the use of a traditional value function approach impossible. Instead, use has been made of stochastic multi-criteria acceptability analysis (SMAA), a family of inverse preference methods based on an analysis of inverse weight spaces (e.g. [17]), and which were briefly introduced in Section 2.2.3. Current SMAA models all treat uncertain attribute evaluations using probability distributions, but can be fairly easily adapted to incorporate simplified uncertainty formats. The integration of simplified uncertainty formats with SMAA is an additional and somewhat secondary aim of the chapter.

The remainder of the chapter is structured as follows. Section 5.2 gives an overview of the SMAA methods. Section 5.3 describes how these models can be adapted for use with different simplified uncertainty formats. Section 5.4, 5.5, and 5.6 describe the three applications. Section 5.7 provides some collective reflections on the attempts to apply simplified uncertainty formats in practical decision problems.

## **5.2 Stochastic multi-criteria acceptability analysis (SMAA)**

The SMAA family of inverse preference methods are useful in applications where preference information is not precisely known, and typically work by providing information to decision makers about the types of preference information that would lead to the selection of a particular alternative. That is, instead of asking ‘which alternative is best given a particular set of preferences?’, one asks ‘what preferences might make a particular alternative the preferred one?’. A number of SMAA variants have been developed. These differ in terms of the preference model used and thus the type of preference information that

is imprecisely known, but are all based upon Monte Carlo simulation from distributions which govern unknown preference parameters (and attribute evaluations). For example, SMAA variants are available for value function [159, 162], outranking [124], reference point [161, 87], and prospect theory [164] methods. The methods described here are the value-function based SMAA-2 [162], with or without an adjustment for the presence of ordinal criteria (SMAA-O, [160]).

In theory the SMAA methodology allows the involvement of decision makers to lie anywhere between not expressing their preferences at all to a complete assessment of preferences as for conventional prescriptive decision aid. In practice SMAA is often applied in situations where the assessment of information from decision makers is limited. This can occur where it is practically difficult or impossible to explicitly state preference information (for example in antagonistic political situations), where the decision maker is unwilling to expend the time and effort required for assessment, or in the early stages of a decision process where the aim is to narrow down the set of potential alternatives to a smaller shortlist for closer consideration. Some of these conditions are the same as those used in previous chapters to motivate against the use of probability distributions in favour of simplified uncertainty formats. This suggests that, although all current SMAA models treat uncertain attribute evaluations by assessing appropriate probability distributions, they may be made relatively more attractive for the types of decision problems just listed if they are adapted to use simplified uncertainty formats rather than probability distributions.

We once again consider a decision problem consisting of  $I$  alternatives  $\{a_1, a_2, \dots, a_I\}$  evaluated on  $J$  attributes  $\{c_1, c_2, \dots, c_J\}$ . Let  $Z_{ij}$  be a random variable denoting the (possibly stochastic) attribute evaluation of  $a_i$  on  $c_j$ , and  $U$  be a multi-attribute utility function mapping the attribute evaluations of alternative  $a_i$  (denoted  $\mathbf{Z}_i$ ) to a real value using a weight vector  $\mathbf{w}$ . A joint density function  $f_X(\mathbf{Z})$  governs the generation of the  $Z_{ij}$  in the space  $X \subseteq R^{I \times J}$ , and a second joint density function  $g(\mathbf{w})$  governs the generation of imprecise or unknown weights in the weight space  $W$ . Total lack of knowledge is usually represented by a uniform distribution in  $W$ .

Given a particular weight vector  $\mathbf{w}$ , the global utility of each alternative can be computed and a rank ordering of alternatives obtained. SMAA-2 is essentially based on simulating a large number<sup>1</sup> of random weight vectors from  $g(\mathbf{w})$  and observing the proportion and distinguishing features of weight vectors which result in each alternative obtaining a par-

---

<sup>1</sup>The exact number of Monte Carlo iterations that are required to achieve a given level is discussed in [252]. To estimate the acceptability index within  $\xi$  of the true value with 95% confidence, one requires  $1.96^2/4\xi^2$  iterations – so that 10 000 iterations will usually be sufficient to achieve error bounds of 1%. In all applications reported in this chapter computational time is not an issue and so 50 000 iterations have been used.

ticular rank  $r$  (usually the “best” rank,  $r = 1$ ). Let the set of weight vectors that result in alternative  $a_i$  obtaining rank  $r$  be denoted by  $W_i^r$ . SMAA-2 is based on an analysis of these sets of weights using the following descriptive measures:

**Acceptability indices** The rank- $r$  acceptability index  $b_i^r$  measures the proportion of all weights that make alternative  $a_i$  obtain rank  $r$ . The most acceptable alternatives are those with high acceptability indices for the best ranks. The acceptability index is defined by

$$b_i^r = \int_{\mathbf{Z} \in X} f_X(Z) \int_{\mathbf{w} \in W_i^r} g_W(\mathbf{w}) d\mathbf{w} d\mathbf{Z} \quad (5.1)$$

In practice, because SMAA is implemented by generating weights and attribute evaluations randomly using Monte Carlo simulation, the acceptability index  $b_i^r$  is simply the relative proportion of all simulation runs in which  $a_i$  obtains rank  $r$ .

**Central weight vectors** The central weight vector  $\mathbf{w}_i^c$  is defined as the expected center of gravity of the favourable weight space  $W_i^1$ . The central weight vector gives a concise description of the “typical” preferences supporting the selection of a particular alternative  $a_i$ , with the aim of helping decision makers understand how different weights correspond to different choices. The central weight vector is defined by

$$\mathbf{w}_i^c = (1/b_i^1) \int_{\mathbf{Z} \in X} f_X(Z) \int_{\mathbf{w} \in W_i^1} \mathbf{w} g_W(\mathbf{w}) d\mathbf{w} d\mathbf{Z} \quad (5.2)$$

but again in practice this integral would not be evaluated directly; rather the central weight vector would be computed from the empirical averages of all weight vectors supporting the selection of  $a_i$  as the best alternative i.e. the  $j$ -th element of  $\mathbf{w}_i^c$  is the average of all weights for attribute  $c_j$  in  $W_i^1$ .

**Confidence factors** The confidence factor  $p_i^c$  gives the probability that alternative  $a_i$  is the preferred alternative if its central weight vector  $\mathbf{w}_i^c$  is chosen, and gives an indication of the impact of uncertainty in the attribute evaluations (it would not be used in deterministic problems). The confidence factor is defined by

$$p_i^c = \int_{\mathbf{Z}: U_i \geq U_k, \forall k} f_X(Z) d\mathbf{Z} \quad (5.3)$$

but would in practice be computed based on simulations rather than an analytic solution to the integral above.

**Ranges for favourable weights** The minimum (maximum) favourable weight  $w_{ij}^{min}$  ( $w_{ij}^{max}$ ) is the minimum (maximum) weight for attribute  $c_j$  in the favourable weight set  $W_i^1$  (i.e. the minimum (maximum) for which alternative  $a_i$  is still the preferred alternative). These are typically used to show the types of judgements that might

exclude a particular alternative from being chosen. The ranges are defined by

$$w_{ij}^{min} = \min_{\mathbf{w} \in W_i^1} w_j \quad (5.4)$$

$$w_{ij}^{max} = \max_{\mathbf{w} \in W_i^1} w_j \quad (5.5)$$

Minimum (maximum) favourable weight vectors  $\mathbf{w}_i^{min}$  ( $\mathbf{w}_i^{max}$ ) are obtained by collecting together all  $w_{ij}^{min}$  ( $w_{ij}^{max}$ ) for a particular  $a_i$ .

In applications where some of the criteria are measured on ordinal rather than cardinal scales, utilities are generated by randomly generating mappings between the ordinal and cardinal scales. At each iteration, utilities of 1 and 0 are assigned to the most and least favoured levels of the ordinal scale respectively, and  $k - 2$  randomly generated values from  $U[0, 1]$  are assigned (after being appropriately sorted) to the intermediate levels of the ordinal scale. This approach is known as SMAA-O [160].

### 5.3 Integrating simplified uncertainty formats with SMAA

Adapting SMAA models to use simplified uncertainty formats is fairly straightforward. Each uncertain attribute is simply replaced by a number of lower-level attributes which capture the uncertainty in the evaluations on that attribute, using one of the simplified uncertainty formats discussed in previous chapters. This transforms the decision problem into one having the same appearance as a deterministic decision problem<sup>2</sup>, and can be treated by any of the existing SMAA methods with some minor modifications. The precise form of these modifications depends on which uncertainty format (and hence simplified SMAA model) is being used. These are described in turn below, again using the value function-based SMAA-2.

#### 5.3.1 Expected values

Each uncertain evaluation  $Z_{ij}$  is replaced by a single value, its expected value  $E[Z_{ij}]$ . The evaluation of  $a_i$  is given by (3.2) with all  $w_{ij}^R = 0$  i.e.  $U_i^{(ev)} = \sum_{j=1}^J w_j u_j(E[Z_{ij}])$ , with attribute importance weights simulated as for SMAA-2. No random generation of attribute evaluations is required using this (or indeed in any other simplified SMAA) method.

#### 5.3.2 Explicit risk attributes

Each uncertain evaluation  $Z_{ij}$  is replaced by two values: its expected value  $E[Z_{ij}]$  and an explicit risk measure  $R_{ij}$  (for example, the variance of  $Z_{ij}$  or the probability that

---

<sup>2</sup>One implication of replacing each stochastic attribute evaluation with a number of ‘deterministic’ ones is that confidence factors are no longer computed as part of the output shown to the decision maker, since there is no longer any uncertainty around these new attributes and any confidence factor would therefore be equal to one or zero.

$Z_{ij}$  does not exceed some specified target). The evaluation of  $a_i$  is given by (3.2) i.e.  $U_i^{(\text{risk})} = \sum_{j=1}^J w_j u_j(E[Z_{ij}]) - \sum_{j=1}^J w_{ij}^R R_{ij}$ . The simulation of weights can take several forms. If variances are used to measure risk, earlier results suggest that Kirkwood's [150] weights  $w_{ij}^R = (-1/2)w_j u_j''(E[Z_{ij}])$  should be used. This requires the same simulation of  $J$  weights at each iteration as for SMAA-2. If other risk measures (like the probability of poor performance) are used, the most lenient approach – in terms of giving each alternative the maximum chance of obtaining one of the best ranks – is to randomly generate the risk weights  $w_{ij}^R$  and attribute importance weights  $w_j$  together i.e. generate  $2J$  weights at each iteration. This would require that the  $R_{ij}$  on each attribute be scaled beforehand to lie between 0 and 1 (assuming the utility functions  $u_j$  do the same to the expected values). Variances could of course also be treated in this fashion. Any further weight restrictions (e.g. that risk weights should be some multiple of attribute importance weights) can be specified as required.

A potentially useful approach would be to use ordinal assessments of the risk of each attribute i.e. ordinal explicit risk attributes. This would require, for each uncertain attribute, only that the decision maker rank the set of alternatives from most to least risky. It may often be easier for a decision maker to make these ordinal assessment of uncertainty rather than to assess it quantitatively, particularly in the early stages of the decision process or where getting extensive participation from the decision maker is difficult. The SMAA-O method then allows cardinal and ordinal attributes to be treated in the same model.

### 5.3.3 Quantiles

Each uncertain evaluation  $Z_{ij}$  is replaced by  $N_q$  values, where  $N_q$  is the number of quantiles used. The evaluation of  $a_i$  is given by (2.5) i.e.  $U_i^{(\text{quan})} = \sum_{k=1}^{N_q} \sum_{j=1}^J w_{jk} u_{jk}(z_{ij}^{(q_k)})$ . The weights  $w_{jk}$  can be simulated directly or as  $w_{q_k} w_{j|q_k}$ , and other weight restrictions (e.g. equal quantile weights) can be specified as required as part of the simulation. Earlier results suggested that one restriction which may be particularly useful is to restrict quantile weights to be those identified by Keefer and Bodily in [144] i.e. to use quantile weights  $w_{0.05} = w_{0.95} = 0.185$ ,  $w_{0.5} = 0.63$ , and then generate only the  $w_{j|q_k}$ .

### 5.3.4 Fuzzy numbers

Each uncertain evaluation  $Z_{ij}$  is replaced by a fuzzy number, usually triangular or trapezoidal. The fuzzy global evaluation of  $a_i$  is given by

$$\tilde{U}_i = \left[ \sum_{j=1}^J w_j u_j(z_{ij}^{(q_1)}), \sum_{j=1}^J w_j u_j(z_{ij}^{(q_2)}), \sum_{j=1}^J w_j u_j(z_{ij}^{(q_3)}), \sum_{j=1}^J w_j u_j(z_{ij}^{(q_4)}) \right]$$

for trapezoidal fuzzy numbers; for triangular fuzzy numbers  $z_{ij}^{(q_2)} = z_{ij}^{(q_3)}$ . These fuzzy evaluations can be ranked using any of the methods discussed in Chapter 2. At each



iteration, a set of  $J$  attribute importance weights must be generated, as for the conventional SMAA-2. The chosen method for ranking the fuzzy global evaluations may also have weighting parameters that can be randomly generated if necessary. For example, it would be possible to randomly generate values for the right-dominance weight  $\Theta$  used in the fuzzy models of Chapter 3 and 4. Any ranking weights would presumably be simulated independently of the attribute importance weights.

### 5.3.5 Scenarios

Each uncertain evaluation  $Z_{ij}$  is replaced by  $N_s$  values, where  $N_s$  is the number of scenarios used. The evaluation of  $a_i$  is given by (2.7) i.e.  $U_i^{(\text{scen})} = \sum_{k=1}^{N_s} \sum_{j=1}^J w_{jk} u_{jk}(z_{ij}^{(s_k)})$ . As in the case of quantiles, the weights  $w_{jk}$  can be simulated directly or as  $w_{s_k} w_{j|s_k}$ , with other weight restrictions (e.g. equal scenario weights) specified as required.

## 5.4 Decision support for evaluating electricity saving options

### 5.4.1 Background to problem

This section describes a project undertaken by a research unit at the Energy Research Center (ERC) at the University of Cape Town, and commissioned by the South African national electricity supplier Eskom<sup>3</sup>. The project is part of Eskom’s demand-side management (DSM) program. Its primary objective is to develop tools which empower Eskom’s residential customers to make more informed electricity consumption choices, in particular choices on how to reduce their consumption. The tools aim at helping energy users to be (a) aware of the electricity saving options available to them; (b) able to evaluate the options available; (c) actively involved in decision making around their own energy-related behaviour.

This section describes a small part of this ongoing project: the design of decision support for addressing aim (a) above i.e. creating awareness about electricity saving options. The intention of the ERC research group is to create awareness (at least partially) through the development of an information campaign making households aware of a limited number of “best ways to save electricity”. This would be distributed through print, internet, and other channels. Limiting the number of presented options is considered essential in light of the enormous number of possible ways to reduce electricity – some initial screening is required to avoid overloading users with information. The current section describes a pilot study constructing a methodology which performs this screening.

---

<sup>3</sup>Eskom is a parastatal organisation responsible for the generation of approximately 95% of the electricity used in South Africa and 45% of the electricity used in Africa. It is regulated by the National Energy Regulator of South Africa under the Electricity Regulation Act and by the National Nuclear Regulator in terms of the National Nuclear Regulatory Act. See <http://www.eskom.co.za/live/index.php>.

## 5.4.2 Problem structuring

### Research objectives

This application follows the distinction drawn in the action-research literature (e.g. [40]) between action aims and research aims. The primary action aim of (this part of) the project is to identify a shortlist of promising electricity savings options to be included on a media insert. It should be possible to customise this shortlist to certain types of users i.e. in the sense of a ‘targeted’ marketing campaign (e.g. [154]). Following similar applications of simplified MCDA models in [197], the research aims here are (a) to establish whether a simplified SMAA model could be employed in a practical decision problem; (b) to identify any difficulties experienced during the application, and (c) to establish whether the additional decision support provided by incorporating assessments of uncertainty was perceived by decision makers to be useful.

### Stakeholder involvement

Ultimate responsibility for formulating electricity saving policy lies with Eskom. Their decisions are based on multiple sources of information collected from external sources – predominantly previous literature and studies conducted by consulting groups such as the ERC. For the purposes of this application, the stakeholder and source of information is a group of three energy researchers at the ERC who are primarily responsible for the project. Inputs to the SMAA models have been provided by these researchers but some of these inputs are based on prior estimates provided by another group of energy researchers at Sustainable Energy Africa (SEA, who often collaborate with the ERC and in this instance provided the data to them). Feedback on the modelling process and model outputs are also provided by those at the ERC. Since the ultimate aim of the media campaign is to reduce household energy consumption, the households are themselves stakeholders in the process, but are only represented in the decision process by members of the ERC group with specialist knowledge of household energy consumption.

### Constructing attributes and alternatives

The set of attributes and alternatives to be used was essentially given by the earlier work done by SEA. They had evaluated 134 electricity saving options on three criteria: electricity savings, capital cost, and the ease of making the change implied by the new technology or behaviour. There is substantial support for these attributes in previous literature around household energy use (e.g. [6, 112, 113]), and these were confirmed in a workshop session with researchers at the ERC.

Measurable attributes had already been defined for each of the criteria. Electricity savings are measured in the number of kilowatt hours saved per month by the introduction of an

electricity saving option. This calculation is based on estimated monthly electricity use before and after the introduction of an alternative. The calculation of prior electricity use usually requires some assumption about the technology/behaviour that a household might use to perform the same function. For example, in the evaluation of a gas oven an electric oven is used as a basis for comparison. The calculation of (before or after) electricity usage is based on multiplying the frequency of usage with the amount of electricity required per “use”. Secondary information sources have been used to indicate (a) the electricity required “per use” prior to the introduction of an alternative, and (b) the percentage of the previous electricity consumption saved by an alternative.

Capital cost is a monetary cost (measured in Rands) of implementing the different alternatives. Where possible this information has been gathered from various distributors. Ease of change had originally been defined using a simple subjective rating scale (a three-point Likert scale: “easy”, “medium”, and “difficult”) but during the assessment of uncertainty a feeling arose that greater differentiation was needed and this scale was expanded to a seven-point constructed scale using the descriptions shown in Table 5.1.

Value	Description
1	Complete change in behaviour: substantial changes to important aspects of life. Large effort required. Adjustment takes a long time with a significant chance that one could not adjust to the change at all.
2	Large changes required: very substantial effort required. Would take a number of months to adjust to the change. All friends would notice a difference in behaviour.
3	Fairly large changes required: significant effort required. Would almost certainly take some weeks to adjust to the change. Most friends would notice a difference in behaviour.
4	Moderate changes required: change will take some effort. May take a short time to adjust to the change. Some friends would notice a difference in behaviour.
5	Minor changes required: still small in number and require little real effort to implement. Would take no time to adjust to the change. Good chance that close family and friends would not notice the difference in behaviour.
6	Very minor changes: Small in number and require only very minor changes to behaviour. Would take no time to adjust to the change. Even close family and friends would probably not notice the difference in behaviour.
7	Absolutely no change required at all to habits; no difference between behaviour before and after.

Table 5.1: Constructed scale used to measure the “ease of use” criterion in the evaluation of electricity saving options.

SEA had divided the 134 electricity saving options into 9 categories: water heating; lighting; cooking; fridge and freezer; heating, ventilation and cooling; laundry and dishes; renewable energy; standby (essentially, using appliances when needed); and others. Because the current analysis was intended as a pilot study, a set of 8 of the cooking options were chosen as alternatives. These alternatives, which contain new technologies as well as changes to behaviour, are:

1. Buy a solar cooker.
2. Buy a gas stove/oven.
3. Buy a “hotbox” (used for cooking rice, stews, etc.)
4. Use a pressure cooker for food that takes a long time to cook.
5. Use a microwave for cooking, instead of an oven.
6. Ensure that the size of the pot matches the size of the stove plate.
7. Keep oven reflectors clean.
8. Boil only as much water as needed when using the kettle.

As part of their analysis, SEA had assessed the expected performance of each alternative on the three attributes. These mean evaluations are based on a mixture of prior research and literature (particularly for evaluations of electricity savings), market research (for costs of new technologies), and subjective assessment (for ease of use). The mean evaluations were presented on the spreadsheet constructed by SEA, an extract of which is shown in Figure 5.1. Uncertainty around the attribute evaluations is not formally assessed, although in a small proportion of cases (particularly relating to cost) ranges of possible values have been recorded in supplementary notes to the spreadsheet.

### **Uncertainty representation**

Uncertainty exists around the evaluation of all three of the attributes. Previous studies give useful indications of expected electricity consumption requirements and savings, but these typically depend on the implementation of the alternative (e.g. hotboxes) or have only been partially studied and remain the subject of discussion (e.g. clean oven reflectors). Frequencies of usage are also subject to substantial variation both between and within households. Uncertainty (here perhaps more accurately labelled ‘imprecision’) exists around costs for some alternatives because these are described only in broad terms and do not specify the exact make (e.g. of gas oven) or distributor, both of which will

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	These values can be modified for interest or when new data becomes available (e.g. differing prices)																		
2	Intervention	KW	Hrs use/day	KWh/month	Capital Cost (incl VAT)	Compare with	KW	Hrs use/day	KWh/month	% Elec Savings	Additional notes	Elec saving	Cost	Ease					
3	Solar cooker	0.0	0.50	0.0	R 2,225.00	Electric stove/oven	2.0	0.50	30.4	100%	Price range: R450-R4000. Average value of R2225 used.	Large	High	Medium					
4	Gas stove/oven	0.0	0.50	0.0	R 4,500.00	Electric stove/oven	2.0	0.50	30.4	100%	Savings of 563Wh per meal (45 minutes of simmer time on a stove plate at 750W)	Large	High	Medium					
5	"Hotbox" used for cooking rice, stew, etc.	1.0	0.20	6.1	R 203.00	Cooking on stove for full duration	1.0	0.50	15.2	60%	Price range: R155-R250. Average price of R203 used.	Large	Low	Easy					
6	Use a pressure cooker for food that takes a long time to cook	1.0	0.25	7.6	R 1,500.00	Use of normal pot	1.0	0.50	15.2	50%		Medium	Medium	Easy					
7	Behaviour: Ensure that the size of the pot matches the size of the stove plate	1.0	0.38	11.4	R 0.00	Pot is much smaller or larger than stove plate	1.0	0.50	15.2	25%		Small	Low	Easy					
8	Behaviour: use a microwave for cooking, instead of an oven	1.3	0.50	19.8	R 0.00	Electric stove/oven	2.0	0.50	30.4	35%	Microwave: 1.3kW Oven (bake element): 1.9kW Oven (grill element): 2.2kW Oven (average): 2kW	Medium	Low	Easy					
9	Behaviour: keep oven reflectors clean	2.0	0.50	29.7	R 0.00	Do not maintain oven	2.0	0.50	30.4			Small	Low	Easy					
10	Behaviour: boil only as much water as needed	2.4	0.01	0.7	R 0.00	Boil an entire kettle full of water unnecessarily	2.4	0.07	5.0	86%		Large	Low	Easy					
11	Behaviour: when using the kettle																		

Figure 5.1: Extract from spreadsheet used by decision makers to inform selection of electricity saving options.

affect costs. More precise definition of alternatives (in terms of makes, etc.) is not considered to be appropriate for the media campaign. Costs for many of the behaviour-type alternatives require no capital outlay, and are deterministic. Evaluations of ease of change are subject to both between household (some households may find one alternative harder to adapt to than another household) and within household (since households cannot be sure in advance of how difficult a new alternative may be to implement and adapt to).

A choice of uncertainty format was made based on a discussion involving the facilitator and ERC research group. Because of constraints on time and effort the group preferred to use only a single uncertainty format, so that it will not be possible to directly compare uncertainty formats (this is true in all three applications). Probability distributions were felt to be too time- and effort-consuming to construct, particularly bearing in mind that the application might ultimately be applied to a large number of alternatives. There were no clear indications of a preference in the group for any particular simplified uncertainty format. After a brief discussion around the results of the previous chapters advocating the use of quantiles in cases as a good ‘default’ option, the group were satisfied to assess uncertainty around the mean evaluations using the 5% and 95% quantiles for each attribute evaluation.

The assessment of the quantiles was performed by one member of the ERC research group who carried out the assessment on his own following a brief meeting to explain the task. Facilitated assessment was not used because the researcher was well-trained in statistics (a postgraduate degree) and had previous experience in the assessment of quantiles. After some discussion it was decided to present results as “worst-case”, “mean/median”, and “best-case” rather than referring to the specific quantile involved. This was done because the same extreme quantile (say, the 5% quantile) is ‘good’ for some attributes (cost) and ‘bad’ for others (electricity saved, ease of change), which the researcher felt may cause some confusion later on. The full value tree for the problem is shown in Figure 5.2 using a format which shows performance at different quantiles as higher-level criteria, as described in Section 2.4.3.

In some cases the quantile assessments were assisted by ranges of possible costs provided in an accompanying note to the original evaluations (e.g. costs of solar cookers and hotboxes; see Figure 5.1). A second meeting was held following completion of the task to discuss the quantile assessments. Means and medians were felt to be reasonably similar (i.e. symmetric distributions) and the original means estimated by SEA were therefore used as central estimates. The final evaluations are given in Table 5.2. The researcher who assessed the quantiles found the learning associated with the assessment process useful as it forced him to think of the underlying factors influencing in particular electricity savings.

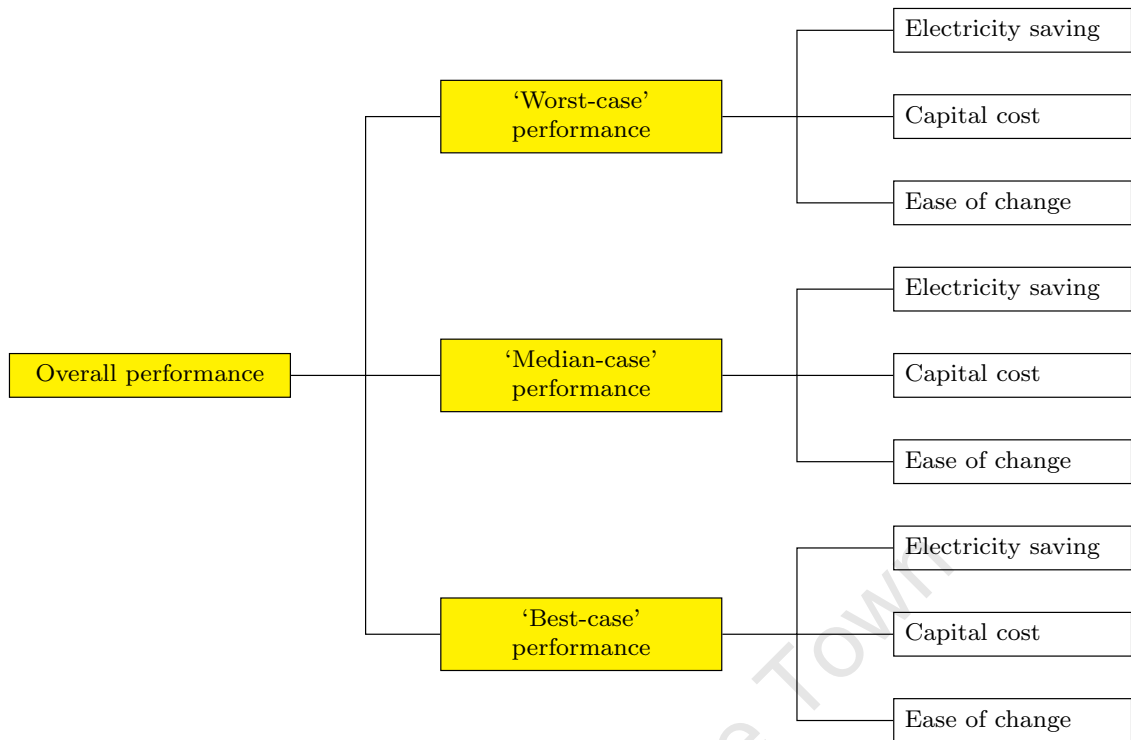


Figure 5.2: Value tree for selecting promising electricity savings options.

The assessment of precise quantiles was felt to be difficult at the extremes, and evaluations there were self-reported as “not quite best” and “not quite worst” cases.

### 5.4.3 Results

Results were obtained by applying a SMAA-O model to accommodate the ordinal ease of change attribute. Utility functions for the two cardinal attributes were randomly generated as per the SMAA-2 model to be convex below a reference point and concave above it, according to general prospect theory [139] principles. Costs were defined as negative profits so that utility functions are increasing in all three attributes. Reference levels

Alternative	Elec savings			Capital cost			Ease of change		
	Wor	Med	Best	Wor	Med	Best	Wor	Med	Best
Solar cooker	2	10	20	4500	2225	300	3	5	7
Gas stove/ovens	30.4	30.4	30.4	20000	4500	1000	2	2	4
Hotbox	2	9.1	25	250	200	150	2	3	4
Use pressure cooker	5	7.6	12	1200	600	240	3	5	6
Pot matches plate	1	3.8	5	0	0	0	1	2	3
Microwave	8	10.646	13	875	250	150	4	5	7
Keep oven reflectors clean	0.5	0.75	1	0	0	0	2	3	4
Boil water as needed	4	11.3	14	0	0	0	2	3	4

Table 5.2: Decision table for evaluation of electricity saving options

reflect the status quo of no electricity savings and zero cost, so that utility functions for the electricity savings attribute(s) are generated to be between linear and moderately concave ( $\tau = \lambda = 0$  and  $0 \leq \beta \leq 4$  using the formulation in (3.7)); and utility functions for the capital cost attribute(s) are generated to be between linear and moderately convex ( $\tau = \lambda = 1$  and  $0 \leq \alpha \leq 4$ ). Utility functions for the ordinal ease of change attribute are simulated by randomly generating sets of cardinal values that are consistent with the known ordinal values as per SMAA-O. Three models were used here to show the impact of uncertainty representation and quantile weight generation, although typically (i.e. outside of the current ‘action research’ context) not all of these would be shown to the decision maker.

**Model 1** Uses the mean/median evaluations only.

**Model 2** Uses the full set of evaluations in Table 5.2, with Keefer-Bodily quantile weights.

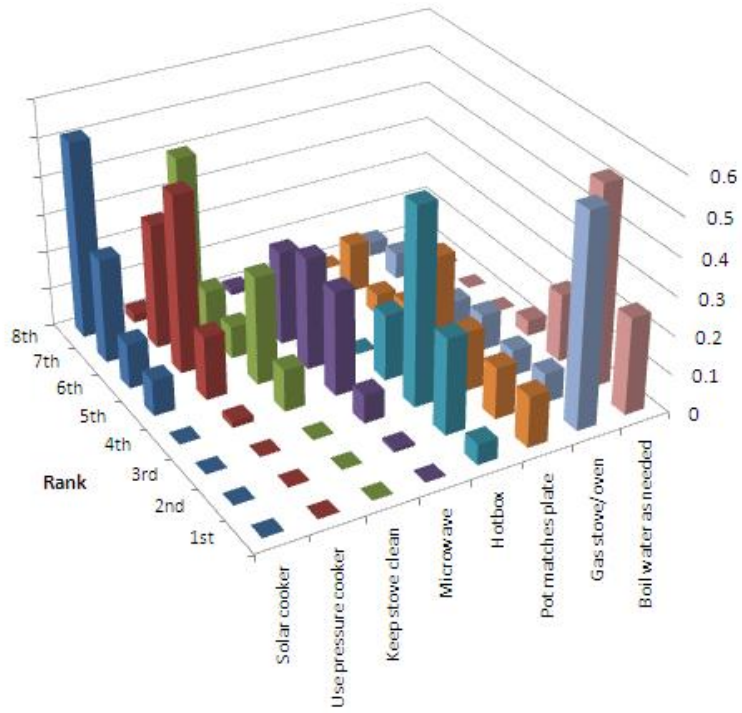
**Model 3** Uses the full set of evaluations in Table 5.2, with quantile weights allowed to vary freely.

Model 1 is used so that the impact of using expected values only can be evaluated. Model 2 is shown to evaluate the sensitivity of results to restrictions on the quantile weights (no restrictions are made on the attribute importance weights), and Model 3 implements the least restrictive approach. All models employ a single locally-defined (between observed best and worst performance evaluations) utility function for each attribute. All three models were presented to energy researchers at the ERC for feedback. The results of each of these models are given below with a summary of the feedback provided.

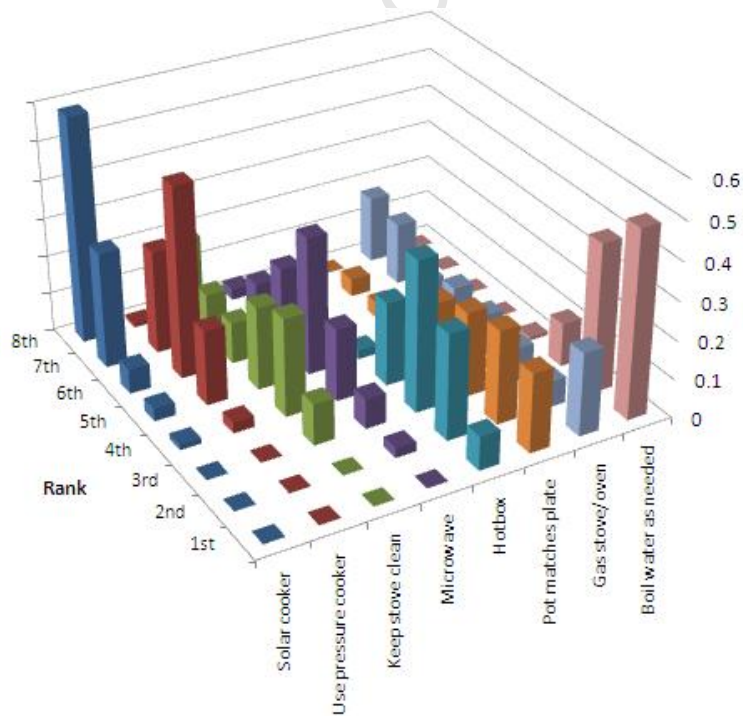
Figure 5.3 displays the rank acceptability indices for each alternative, showing the share of the different preferences i.e. weights, that support an alternative for a particular rank 1 (best) through 8 (worst). Acceptability indices in Figure 5.3(a) are calculated based on quantiles with Keefer-Bodily quantile weights (Model 2). Those in Figure 5.3(b) are calculated based on quantiles with unrestricted quantile weights (Model 3). Figures 5.3(c), (d), and (e) show acceptability indices based only on attributes in the ‘worst-case’, ‘median-case’, and ‘best-case’ part of the value tree respectively. Note that the ‘median-case’ Model 3 results are effectively those of Model 1 (which uses only median evaluations).

The most promising alternatives are those with larger (at least non-zero) acceptability indices for the best ranks, which appear towards the bottom-right corner of each of the plots in Figure 5.3. There was reasonably strong agreement between the models on the most promising alternatives. All models identify gas stoves/ovens, matching pots to stove plate sizes, and boiling water only as needed as alternatives that potentially appear first in a preference order. The sum of the rank-one acceptability indices for these three alternatives



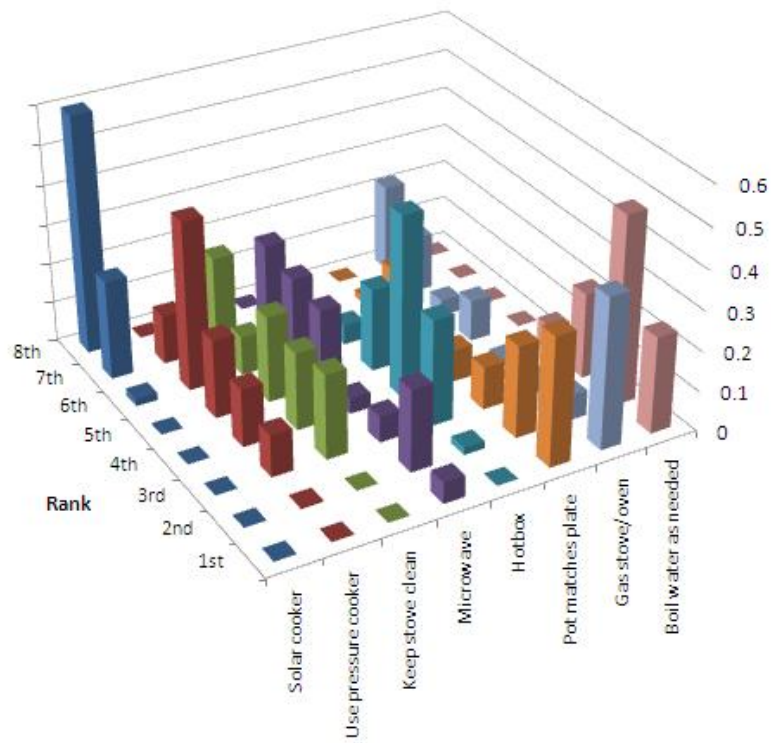


(a) Model 2

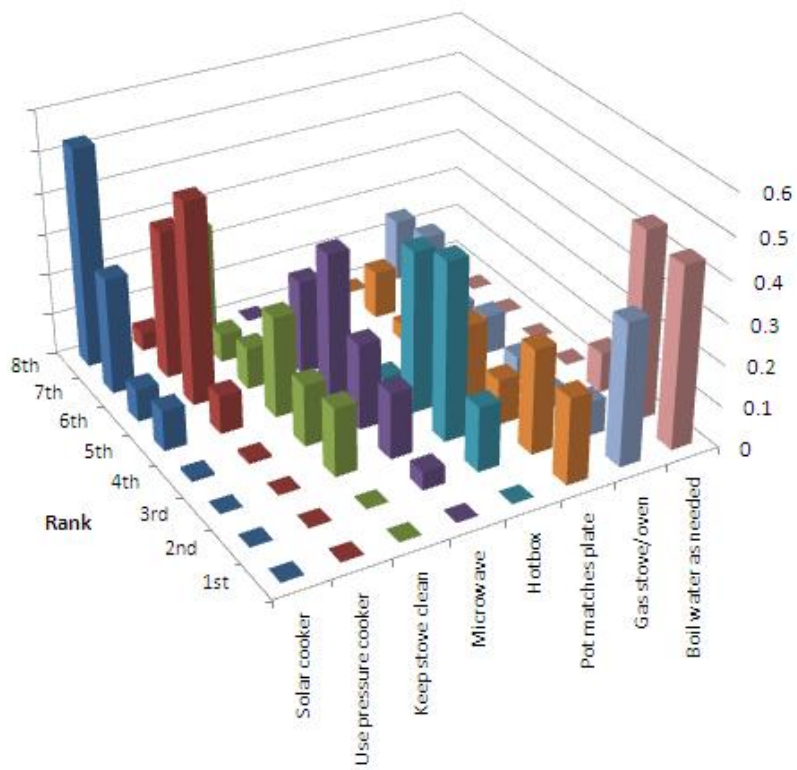


(b) Model 3

Figure 5.3: Acceptability indices for electricity savings options obtained with an adapted SMAA-O model.

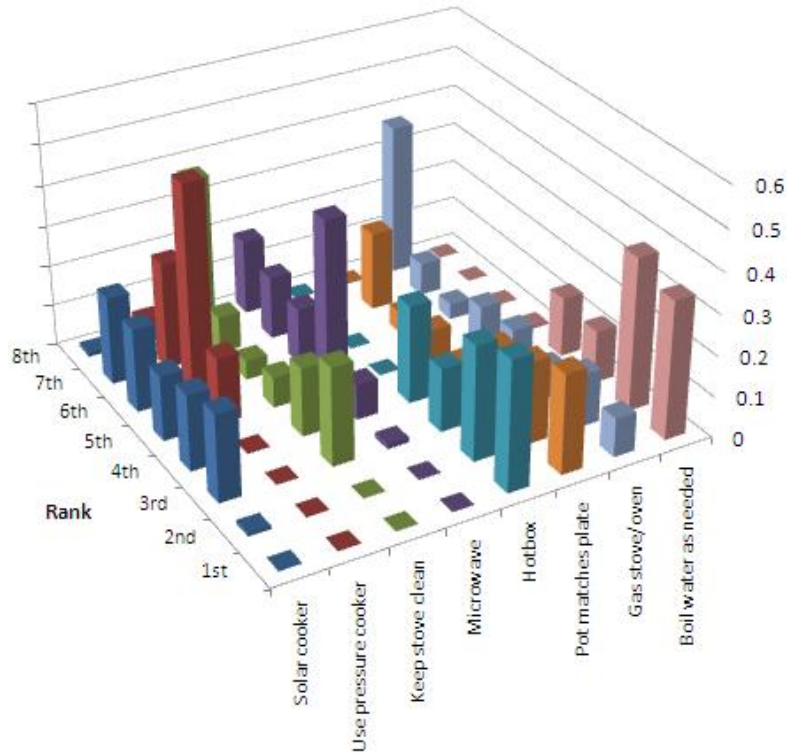


(c) Model 3, worst cases only



(d) Model 3, median cases only

Figure 5.3: Acceptability indices for electricity savings options obtained with an adapted SMAA-O model.



(e) Model 3, best cases only

Figure 5.3: Acceptability indices for electricity savings options obtained with an adapted SMAA-O model.

is (in Model 3) 0.91, signifying that 91% of all possible sets of weights lead to the selection of one of these alternatives. Hotboxes and microwaves also show substantial acceptability indices for second and third ranks, with hotboxes in particular becoming relatively more favourable in the models using quantiles (Models 2 and 3). The other three alternatives – keeping oven reflectors clean, pressure cookers, and solar cookers – perform noticeably worse and tend to occupy lower ranks. The main difference between the results of Model 2 and Model 3 is that the two alternatives with the highest rank-one acceptabilities – gas stoves/ovens and boiling water as needed – swap ranks (in Model 2, gas stoves/ovens have the larger rank-one acceptability; in Model 3, boiling water as needed). Because the aim of the analysis is to identify a shortlist of promising alternatives the reversal was not considered to be particularly important, but its cause is that boiling water as needed is favoured by a greater range of quantile weights, in particular those which place relatively more weight on the ‘best case’ quantile than the Keefer-Bodily weights do (as shown by the relatively better performance of boiling water as needed in Figure 5.3(e)).

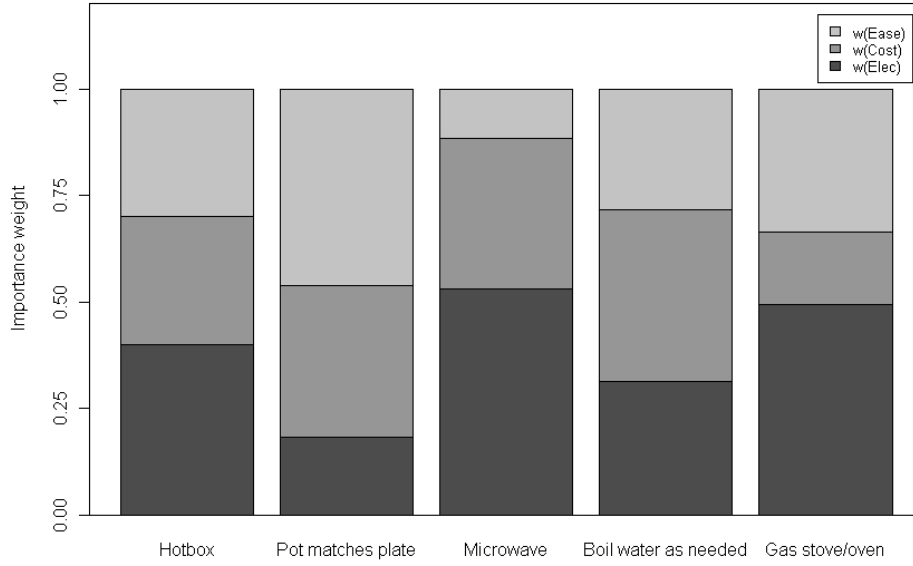
In an informal feedback discussion, the group indicated that they found the acceptability results easy to understand and interpret. They identified the ability of the SMAA model to identify a diverse set of promising alternatives with minimal additional information as

its main benefit, and it became fairly clear during the analysis that a promising shortlist of alternatives should include gas stoves/ovens, matching pots to stove plate sizes, boiling water only as needed, and hotboxes. The selection of promising alternatives was experienced as fairly easy and was based on (a) the acceptability indices for the best ranks and (b) the robustness of the rank of an alternative to changes in preferences. The latter criterion was felt to be of importance given the aim of promoting alternatives that would be viewed as acceptable by as many of the target audience as possible. Hotboxes and boiling water as needed were identified as the most robust of the promising options, and gas stoves/ovens and matching pot and plate size as relatively sensitive to preferences. The poor robustness of microwave ovens also played an important role in the group's decision to exclude this alternative from the shortlist. The extra (within-quantile) acceptability information provided by Model 3 was considered useful in better understanding the performance of the alternatives but in this instance did not affect the selection of alternatives.

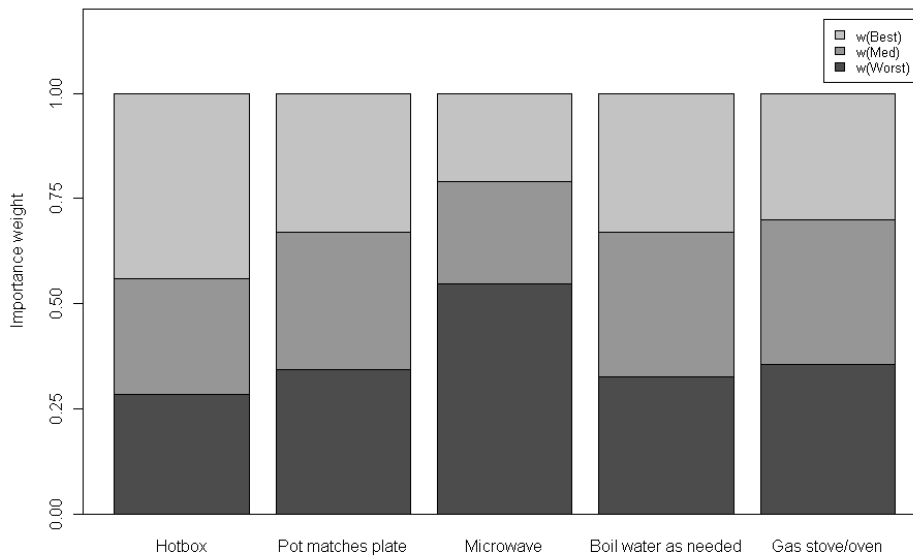
Figure 5.4 displays the central weights for each alternative with a non-zero rank-one acceptability index – these show the typical (average) weights that make each alternative most preferred. A large shaded area for an attribute indicates that a greater weight is allocated to that attribute in the central weight vector. Figures 5.4(a), (b), and (c) show the attribute importance weights of efficiency attributes in the 'worst-case', 'median-case' (i.e. Model 1), and 'best-case' part of the value tree respectively. Figures 5.4(d) and (e) show attribute importance and quantile weights respectively; these are calculated over the entire value tree. Figure 5.4(f) shows a joint presentation of attribute importance and quantile weights originally shown to the ERC group. All figures are based upon Model 3 – very similar attribute importance weights, both over all quantiles and within each quantile (i.e. those shown in Figures 5.4(a), (d), (e), and (f)) were obtained using Model 2. These are therefore not shown here.

The central weight vectors shown in Figure 5.4 provide insight into the reasons why certain alternatives may be preferred. Figure 5.4(a) displays the central weights for the three attributes, and shows that:

- The use of a hotbox tends to be preferred by decision makers who place roughly equal weights on the three attributes, with slightly more weight being placed on electricity savings.
- Matching pots to plate sizes tends to be preferred by decision makers with relatively large weights on ease of change, and small weights on electricity savings.
- The use of a microwave tends to be preferred by decision makers with relatively large weights on electricity savings, and small weights on ease of change.

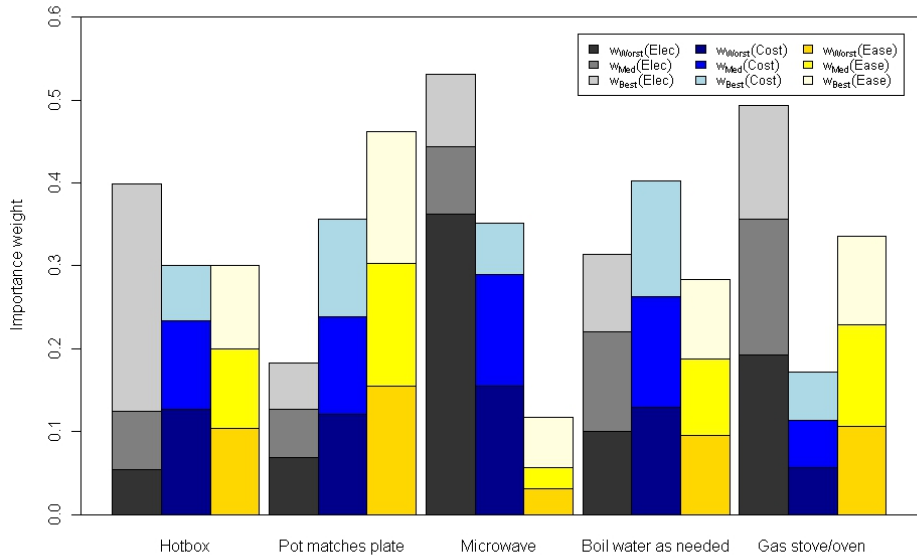


(a) Attribute weights (all quantiles)

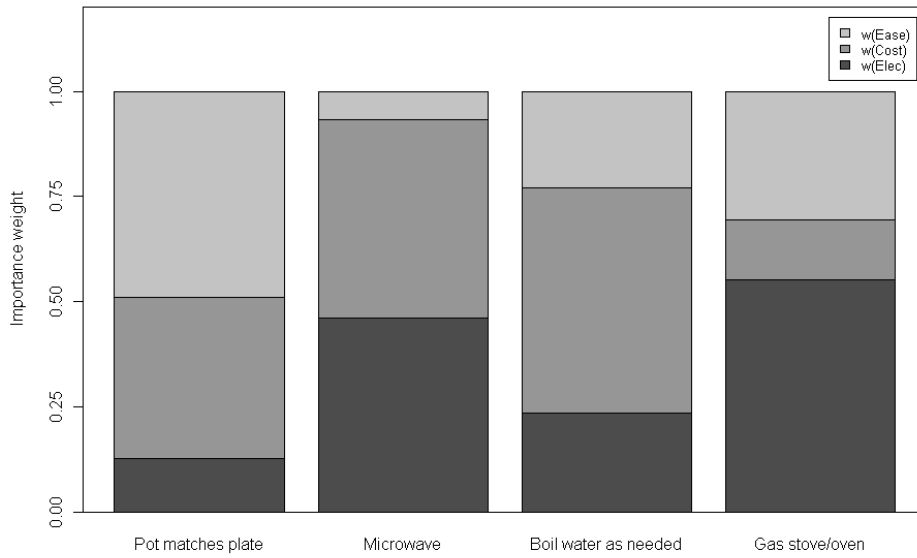


(b) Quantile weights (all attributes)

Figure 5.4: Central weight vectors for electricity savings options obtained with an adapted SMAA-O model. All figures are based upon Model 3.

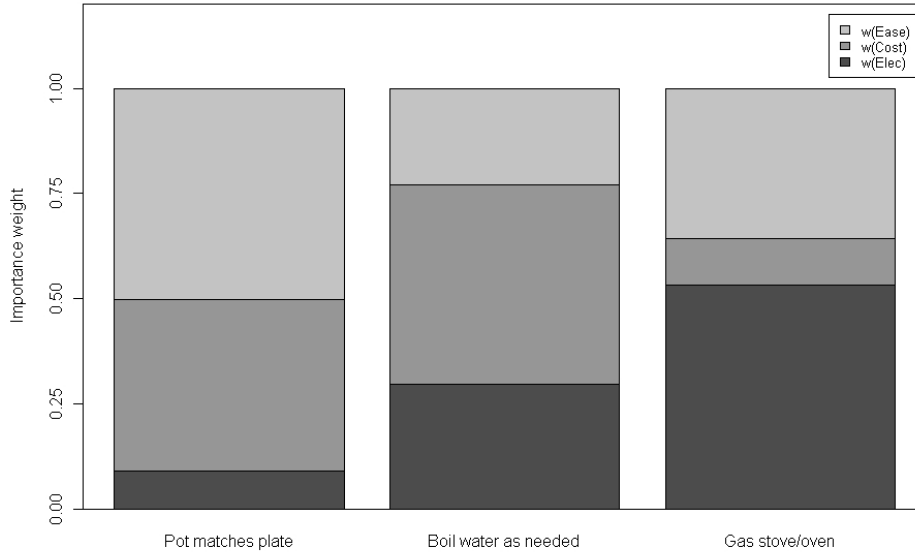


(c) Joint weights

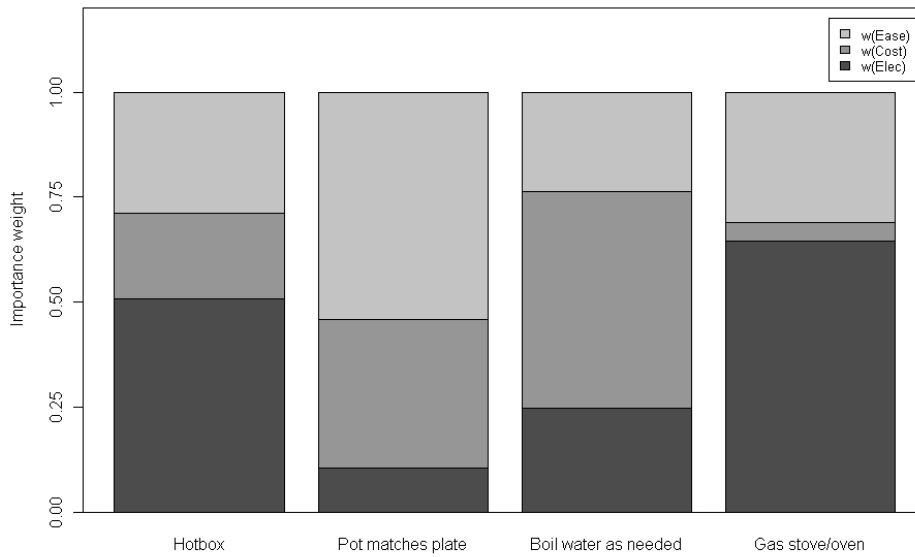


(d) Worst cases only

Figure 5.4: Central weight vectors for electricity savings options obtained with an adapted SMAA-O model. All figures are based upon Model 3.



(e) Median cases only



(f) Best cases only

Figure 5.4: Central weight vectors for electricity savings options obtained with an adapted SMAA-O model. All figures are based upon Model 3.

- Boiling water only as needed tends to be the preferred action of decision makers who place roughly equal weights on the three attributes (as for hotboxes), but with slightly more weight being placed on cost.
- The use of gas stoves/ovens tends to be the preferred action of decision makers with relatively small weights on cost and larger weights on electricity savings.

The central attribute weights were easily understood and interpreted by the group, and confirmed “common sense” descriptions of the alternatives. Note that Figure 5.4(a) shows that if only medians are used (Model 1), then no descriptions are provided for hotboxes or microwaves because their rank-one acceptabilities are in that case zero. In this application the addition of other quantiles therefore had a substantive impact on the information that could be displayed. It was proposed that the acceptability indices in Figure 5.4(a) could also be used to tailor intervention recommendations (the shortlist of suggested alternatives) to different decision maker “typologies”. These typologies are currently expressed in terms of importance weights on the three attributes (as in the above list), but ultimately these may be linked to other characteristics – for example low-income households placing greater importance on cost. This possibility for customising awareness programs was identified as a second major benefit of the analysis.

Figure 5.4(b) shows the central quantile weights for Model 3. Approximately equal central quantile weights are observed for three alternatives: matching pot and stove plate size, boiling water as needed, and gas stoves/ovens. The equal weights, in conjunction with the relatively large rank-one acceptability indices for these alternatives, suggest that a wide range of quantile weights support each of them (rather than quantile weights *having* to be equal). The other two alternatives show some patterns. The use of a hotbox tends to be preferred if more weight is placed on the ‘best-case’ evaluations, and the use of a microwave tends to be preferred if more weight is placed on the ‘worst-case’ evaluations (both under the condition that attribute weights are also allocated as indicated in the list above).

The group felt that the central quantile weights were moderately useful. Although the hotbox had already been identified as a promising alternative, knowing that its attractiveness depends quite heavily on being applied ‘ideally’ was felt to be a practically useful insight. In this case this had already been suggested by evaluating the acceptability indices within each quantile; the central weights provide a second, more direct indication. The information provided by the central quantile weights was considered of relatively minor importance when compared with that provided by the central attribute weights, and it noted that Model 2 (with constant Keefer-Bodily weights) provided essentially the same conclusions. However the group accepted that results would not necessarily be so similar



in other cases, and given the greater flexibility of Model 3 and the limited additional effort in interpreting its results, it was chosen as the preferred model for future work. Finally, it should be noted that a joint presentation of central attribute and quantile weights, as shown in Figure 5.4(c), was initially attempted. The group found the interpretation of these weights relatively difficult and preferred the aggregated results shown in Figure 5.4(a) and (b). The difficulty seemed to lie with understanding trade-offs involving differences in both attributes and quantiles (e.g. between worst-case cost and best-case electricity saving). This may be expected to be as or more problematic for the elicitation of attribute and quantile weights.

#### **5.4.4 Follow-up**

The ERC project is in its early stages but the group is satisfied that the SMAA approach described in this section can contribute to the design of targeted media around electricity savings options, by supporting the decision of which options to include in the media. Similar analyses to the one performed here for cooking interventions are to be carried out for other categories of energy use (e.g. cooking, lighting, heating) in the months to come. The choice of media platform(s) and the design of the media is scheduled for 2011. In the longer term the project aims to evaluate the impact of targeted campaigns on awareness of energy-related issues and making consumers actively involved in decision making around their own energy-related behaviour, but it is not yet possible to assess whether the shortlists of alternatives provided by the SMAA model are effective in helping to achieve these goals.

## **5.5 Decision support for evaluating slot machines**

### **5.5.1 Background to problem**

This application arose from a project undertaken for a large casino in the Western Cape province of South Africa. The focus of the project was on constructing simulation models of some of the machines on the casino floor, but during this task it became clear that different makes of slot machines held by the casino differed quite substantially in terms of the distributions of their payouts. Literature around gambling behaviour suggests that some distributional summary measures, particularly expected values and probabilities of payout, are linked to psychological factors affecting gambling behaviour, and thus might be used as measurable attributes for evaluating the desirability of a machine to a particular type of player. A secondary study was carried out with the aim of describing different slot machine varieties at the casino in terms of these attributes, and in terms of the types of players (i.e. attitudes towards playing) that the casino is currently catering for.

## Basics of slot machine play

The main ingredient of a slot machine is a set of reels, usually five, each of which can be thought of as a list arranged in a circular configuration. When the machine is played, the reels are “spun” and a position on each reel is chosen at random. The symbol in the selected position on each reel, as well as the symbols in the positions adjacent to the selected position on the reel, are displayed graphically to the player in a  $3 \times 5$  grid. Payouts are made when the same symbol appears along certain winning lines across the grid. A winning line is a sequence of cells appearing from left to right in the grid. For example, the first three winning lines in all slot machine games run horizontally across the cells in the middle, top and bottom rows of the grid respectively. A selection of typical winning lines is given in Figure 5.5. Some modern machines can be played on up to 25 winning lines.

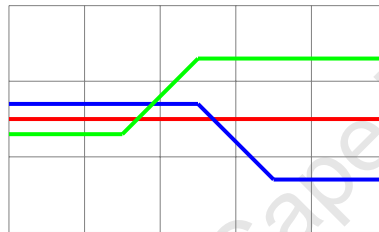


Figure 5.5: Illustration of three typical ‘winning lines’ on a slot machine.

Before each play, players choose the number of winning lines they would like to make use of and the amount they would like to bet per winning line, so that the total amount bet per play is given by the product of the two. For a payout to be won on a winning line the same symbol must appear in the first  $P$  reels from left to right along this line, with the payout usually only being non-zero for  $P \geq 3$  and increasing with  $P$ . Following each play, the winnings on each of the played lines are summed and added to the player’s winnings. Different symbols offer different payouts, with the symbols that appear most rarely on the reels generally giving the biggest payouts when they do appear.

### 5.5.2 Problem structuring

#### Research objectives

The ultimate objective of the casino is to provide a portfolio of slot machine that are popular with their customers. This involves identifying attributes that are important to different segments of slot machine players, and ensuring that at least one of the portfolio of machines held by the casino addresses the desires of each segment. Because no formal evaluation of slot machines had been previously attempted by the casino, the primary action aim of the project is simply to describe the slot machines in a systematic and formalised way, and to evaluate any market segments (i.e. attitudes towards playing)

that are not catered for by the portfolio of machines currently held at the casino. This places the analysis within what Roy [222] identifies as the “description” problematique. A secondary action objective is to identify any clearly inferior i.e. dominated, make of slot machine. Research aims are: (a) to establish whether a simplified SMAA model could be employed in a practical decision problem; (b) to identify any difficulties experienced during the application, and (c) to establish whether the additional decision support provided by incorporating assessments of uncertainty was perceived by decision makers to be useful.

### **Stakeholder involvement**

Partly due to the size of the casino company involved, the analysis was conducted at a level quite far removed from any final decision making authority. The construction of a decision analytic framework was commissioned by a statistical consultant working for the casino, who liaised with a line manager in charge of technical operations at the casino. This line manager would then report to more senior management, who had ultimate responsibility for major decisions. The line manager could therefore be seen as one stakeholder group tasked with informing or potentially seeking to influence senior management, and the consultant as another stakeholder tasked with the provision of information to the line manager. However, the client in this case was very much the statistical consultant, who commissioned the work and provided input regarding technical details as well as feedback throughout the structuring process. The line manager did not participate directly in any of the modelling.

### **Constructing attributes and alternatives**

Chantal et al. [52] identify two categories of motivations for gambling: gambling for intrinsic rewards like excitement, a sense of accomplishment, or to socialise with friends; and gambling to gain external rewards, usually money. The objective of maximising monetary gains (or at least minimising losses) is an obvious one, and can be measured by the long-run expected return from each play. Different machines offer different expected returns – these are routinely provided to casinos by slot machine manufacturers but are typically not disclosed to the public (they are also always negative). Provincial gambling legislation sets the minimum expected return to the gambler (calculated over a continuous 12-month period) at 85% of the amount bet<sup>4</sup>.

Excitement is the most frequent self-reported motivation for regular slot machine playing [86]. Players “tend to prefer more frequent reinforcement rather than larger and less fre-

---

<sup>4</sup>Gambling in the Western Cape is governed by the *Western Cape Gambling and Racing Board Casino Rules* published in Provincial Gazette Extraordinary 5952 (29 November 2002) and subsequently amended in Provincial Gazette Extraordinary 5983 (14 February 2003). The full set of rules is available at <http://www.wcgrb.co.za/>. This particular rule is contained in Section 7.12, point (9).

quent reinforcement” [84], which is supported by evidence that most players play multiple lines with the minimum number of credits [269]. A simple way to operationalise expected entertainment is to use the probability of winning *any* kind of payout on a particular play, termed the ‘win probability’. The motivation behind the use of the win probability is that it is the act of winning some kind of payout that contributes significantly to the excitement of the player – including audiovisual activities such as lights going off, sounds of coins dropping, and so on, which have been shown to be important components of entertainment [179] – and that long periods of inactivity are likely to be construed as boring intervals between the exciting winning plays. Payne [212] has recently provided evidence that overall probabilities of winning or losing are important influences on risky choice in fairly simple (five-outcome) gambles.

Because slot machines can be played on different numbers of lines, and this may affect the performance of a machine in terms of monetary return and excitement, alternatives are defined by combinations of a slot machine type and a number of lines to play. Although a casino has no influence over how many lines are played, a casino may decide to hold a particular type of slot machine because (perhaps only for a specific number of lines played) it is popular with a segment of their customers.

### **Uncertainty representation**

The current application consists of a single uncertain criterion (monetary returns) which is decomposed into a ‘value’ component (expected return) and a ‘risk’ component (the probability of achieving a return greater than zero). The choice of uncertainty format has therefore essentially been made as part of the attribute definitions. Information on expected returns and win probabilities was gathered for five makes of slot machines on the casino floor at the time, by creating simulated versions of the machines using reel information provided to the casino by the machine manufacturers. The machines differ on the number of lines that can be played: Machine *A* can be played on 1, 3, 5, or 9 lines; Machines *B* and *C* on those of *A* as well as 15 lines; and Machines *D* and *E* on those of *B* and *C* as well as 21 and 25. There are therefore 29 alternatives in total. Table 5.3 shows the (negative) expected returns and win probabilities for the set of alternatives – all statistics have had a small constant added to them to preserve the proprietary nature of the information, but this makes no difference to results.

### **5.5.3 Results**

Results were obtained by applying a SMAA-2 model with utility functions for the expected returns generated to be between linear and moderately convex (since they are losses). Lin-

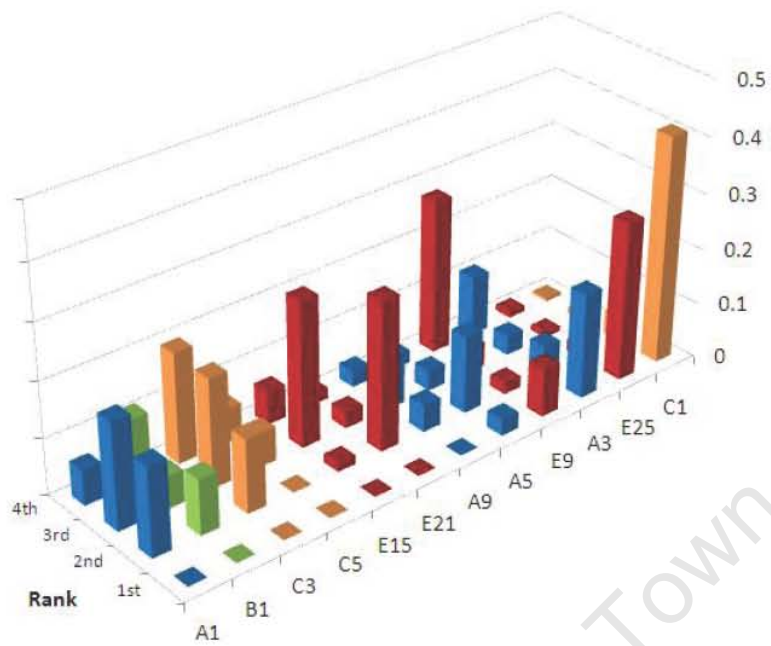
Lines	Slot A		Slot B		Slot C		Slot D		Slot E	
	<i>EV</i>	<i>Pr</i> <sub>0</sub>	<i>EV</i>	<i>Pr</i> <sub>0</sub>	<i>EV</i>	<i>Pr</i> <sub>0</sub>	<i>EV</i>	<i>Pr</i> <sub>0</sub>	<i>EV</i>	<i>Pr</i> <sub>0</sub>
1	-0.05	0.15	-0.04	0.05	-0.02	0.11	-0.10	0.12	-0.06	0.15
3	-0.09	0.33	-0.08	0.12	-0.05	0.20	-0.24	0.20	-0.14	0.27
5	-0.14	0.39	-0.12	0.18	-0.09	0.27	-0.36	0.27	-0.20	0.38
9	-0.23	0.46	-0.19	0.26	-0.15	0.36	-0.55	0.39	-0.33	0.54
15			-0.28	0.33	-0.24	0.42	-0.73	0.43	-0.48	0.59
21							-0.85	0.47	-0.60	0.63
25							-0.90	0.49	-0.66	0.65

Table 5.3: Attribute evaluations for five slot machines

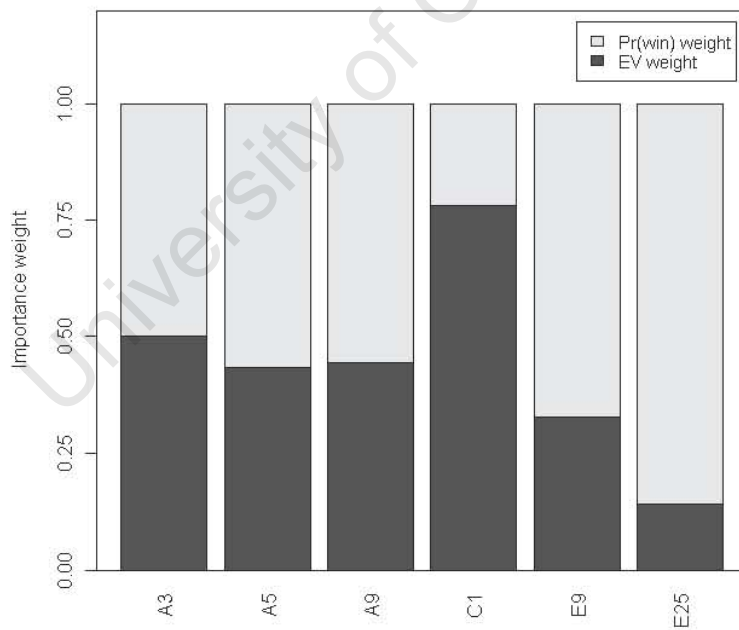
ear utility functions were used for the win probability attribute in accordance with (2.4). In this application there is only a single underlying criterion (monetary return) captured by one ‘value’ attribute (expected return) and one ‘risk’ attribute (win probability). There are therefore only two weights to be generated at each iteration. Figure 5.6 displays the acceptability indices and central weight vectors. Alternatives are referred to by concatenating their machine make (A to E) and the number of lines played (1 to 25).

Three of the five machine types are non-dominated: Machine *A*, *C* and *E*. The alternative with the largest rank-one acceptability index is Machine *C* played on a single line, following by Machine *E* played on 25 lines. Machine *A* contributes three non-dominated alternatives, when played on three, five, or nine (its maximum number of) lines, although the acceptability indices for three-line and five-line plays are small (3% and less than 1% respectively). The only other non-dominated alternative is Machine *E* played on 9 lines. The central weight vectors show that alternative *C*1 caters for players who place considerably more importance on maximising expected return than on maximising entertainment; alternative *E*25 caters for players who do the opposite. Between these extremes Machines *A* and *E* offer some alternatives that provide greater a balance between the two objectives. The following set of descriptions was provided to the decision maker:

- Machine type *A* is the preferred option of a range (22%) of possible player types. It appeals to players who place roughly equal weight on monetary return and entertainment, by offering a balance between these two attributes.
- Machine *B* is one of the machines that is dominated by others. It appears unfavourably because of a poor win probability accompanied by relatively moderate expected returns. For a given expected return, Machine *A* consistently offers greater entertainment value.
- Machine *C* is the preferred option of the widest range (41%) of possible player types. It appeals to players seeking ‘explicit rewards’ i.e. those who place greater



(a)



(b)

Figure 5.6: Results of a SMAA-2 analysis showing, (a) acceptability indices for the first four ranks for all machines having any of the corresponding acceptability indices greater than zero, (b) central importance weights for all machines with rank-one acceptability greater than zero.

importance on monetary return than entertainment. For any number of lines played it offers the best expected return of all the machines, at the expense of a relatively low win probability. It is particularly attractive for a small number of lines played (one or three); beyond that, it is dominated by Machine *A*.

- Machine *D* appears by some way the worst of the modelled machines. It is dominated by Machine *E* for all lines played, by Machine *A* up to 21 lines, and by Machine *C* up to 15 lines.
- Machine *E* is the preferred option of a range (37%) of possible player types. It appeals to players seeking ‘implicit rewards’ i.e. those who place greater importance on entertainment than monetary return. It offers the greatest entertainment if 9 or more lines are played, but at fewer than 9 lines is dominated by Machine *A*.

In an informal feedback discussion the client indicated that both the acceptability indices and central weight vectors were intuitive and easy to interpret. The correspondence between the SMAA results and observed gambling behaviour was felt to create confidence in the method and in the set of attributes used. One expressed concern was that qualitative characteristics of slot machines – particularly the entertainment value of bonus games and the attractiveness of the main “theme” – was not taken into account. The client felt that Machine *B*, for example, had a novel bonus game which might mitigate its relatively poor performance on the quantitative attributes. Because of the difficulty of obtaining subjective evaluations from the management of the casino, a decision was made not to include these qualitative attributes at this stage, although as shown in the previous application it is entirely possible to do so using either a constructed cardinal scale or an ordinal scale in conjunction with SMAA-O.

#### 5.5.4 Follow-up

The client indicated, again in informal discussion, that he found the learning associated with structuring the problem useful as it led to a perspective on slot machines (involving precise statistical definitions of ‘risk’ and ‘return’) which is quite different to the qualitative one usually considered in the industry. This allows for a more objective comparison of machines. The results obtained from the analysis were presented in report form to the line manager of technical operations at the casino, but obtaining any meaningful feedback from casino management on the methodology or results has proved difficult. A general satisfaction with the results was expressed by the line manager, but a strict code of confidentiality among casino employees meant that requests for information about the actual popularities of slot machines on the casino floor were rejected. Soon after the report was completed the line manager resigned from the casino, so that it is doubtful that the SMAA methodology (which would have been driven by the line manager if at

all) is currently being used at the casino. It is therefore unfortunately not possible to evaluate the usefulness of the SMAA methodology to the end user in this case. Two visits to the casino, one made six months after the work was completed and another made after a further six months, revealed that all the slot machine included in the current study were still in use on the casino floor.

## **5.6 Decision support for evaluating brands in consumer markets**

### **5.6.1 Background to problem**

Consumer behaviour and brand-choice models often concern themselves with the construction of simple metrics that can be easily measured in survey questions and shown to be correlated with actual purchase behaviour e.g. overall satisfaction in [208], intention to repurchase in [215], and various forms of ‘commitment’ in [263] and [111]. These metrics are generally gathered from individual survey respondents and aggregated<sup>5</sup> to identify the changing fortunes of brands and to inform marketing strategy. The purpose of the application described in this section is to construct and validate such a measure from survey questions indicating which alternatives possess each of a set of attributes i.e. a traditional matrix of attribute evaluations that may be binary. The project was undertaken for a market research group in South Africa.

#### **Review of relevant brand-choice models**

There are two broad kinds of descriptive brand-choice model. The first uses a single measure as a proxy for the latent value of an alternative. Typical measures include overall satisfaction [208], affective commitment [263], repurchase intention [215], and overall ‘efficacy’ [215]. The popularity of these measures, particularly commercially, derives from the parsimony of their measurement and the lack of a need for parameter estimation. Their ease of construction means that choice predictions can often be made directly from the single measure e.g. select the alternative with the highest overall satisfaction. Other examples of these models can be found in [76] and [148].

The second type of model tends to use multiple lower-level attributes such as marketing-mix variables or specific product or service characteristics such as price, quality, and so on. These models generally provide a finer level of detail into the reasons for choice but also require more effort in estimation. Furthermore, since different aspects may be important to different people, accommodating respondent heterogeneity using random-effects specifications becomes standard [e.g. 234]. This means that any choice prediction requires sub-

---

<sup>5</sup>The way in which responses are aggregated depends on the sampling approach used but is generally a simple (in the case of random sampling) or weighted (stratified sampling) average.



stantial and often technically sophisticated model fitting using some empirically-gathered dependent variable. Examples of these models can be found in [106] and [224].

The market research group in this study desired an approach which did not require any calibration against an empirically-gathered dependent variable. That is, only models that construct a single measure of brand performance from one or more survey questions, without any parameter estimation, were considered. This excludes all so-called “proper” models of the second type<sup>6</sup>. A model of the second type that can and will be considered however, is an “improper” linear model that assigns an equal weight to each attribute. Dawes [74] has shown that predictions obtained from such an improper model can prove both accurate and robust to changes in external circumstance.

### 5.6.2 Problem structuring

#### Research objectives

The aim of the project is to construct a measure of brand performance that can be used in future studies conducted by the market research company. In order for any measure to be used with confidence (by the market research company and its clients), it is important that it is shown to be associated with actual choice behaviour. The validation of the constructed measure against observed behaviour in two ‘pilot’ studies constitutes a second, related aim. Validation is based on two outcomes: customer retention (the continuance of an individual’s relationship with a service or product) and ‘share of wallet’ allocations (the percentage of money allocated by an individual to a particular service or product in a category). By way of comparison, many authors have shown that overall satisfaction has a positive effect on the intention to repurchase [e.g. 37, 158], retention [e.g. 107], and share of wallet [e.g. 148, 215]. Correlations between satisfaction and share of wallet and retention generally lie between 0.2 and 0.4 [e.g. 158, 148, 215], although much smaller correlations between satisfaction and behaviour have also been observed (e.g. correlations of 0.07 in [231] and 0.13 in [111]; also see [263]). Correlations between repurchase intentions and buying behaviour range from fairly weak ( $r = 0.11$  in [231] and  $r = 0.24$  in [158]; also see [198]) to substantial ( $r = 0.43$  in [137],  $r = 0.44$  and  $0.65$  in [215]).

In consultation with the market research company, the following three research hypotheses (to be satisfied by any prospective measure) were formulated:

**Hypothesis 1:** Relative purchase frequency is positively associated with the constructed performance measure, both at the individual and aggregate level.

**Hypothesis 2:** Probabilities of defection are higher from alternatives with relatively lower

---

<sup>6</sup>Proper linear models are those in which the weights of the input variables are estimated so as to fulfil some prespecified optimality condition, which requires a dependent variable.

constructed performance measures.

**Hypothesis 3:** Changes in relative purchase frequencies are positively associated with the constructed performance measure.

### **Stakeholder involvement**

The formulation of marketing strategies will usually be performed by the clients of the market research company. Although the market research company may play some part in the crafting of strategy, their primary responsibility is to provide survey information to their clients to assist in the decision-making process. This is similar to the Eskom/Energy Research Center relationship described in Section 5.4. The primary stakeholder in this application is the market research company, who (in consultation with their clients) construct sets of relevant alternatives and attributes and conducts surveys to collect attribute evaluation information. The actual evaluation of alternatives over attributes is of course performed by survey respondents. Feedback on results were primarily obtained from two directors of the market research company.

### **Constructing and evaluating alternatives and attributes**

The relationship between the constructed brand performance measure and buying behaviour is evaluated by testing the three research hypotheses described above using real-world longitudinal survey data obtained for two different fast-moving consumer goods (FMCG) markets: the toothpaste market in the United Kingdom (17 brands,  $n_1 = 281$ ) and the laundry detergent market in Spain (11 brands,  $n_2 = 361$ ). This data was kindly provided by Dr Jan Hofmeyr at Synovate (also a market research group). In both studies, a team of researchers assigned to the project constructed the set of brands to be included in the study and a list of relevant attributes based on their own knowledge of the particular market. This problem structuring is not facilitated in any decision-analytic sense. Usually (and in both the studies reported here) effectively all brands participating in the market are included; there may be a small number of brands with almost no market presence that are not explicitly included (although an “other” option can be specified for most survey questions). Attribute sets contain a mixture of attributes that are widely used across product categories (e.g. “price” and “quality” attributes) and those that are specific to a particular category. Table 5.4 lists the attributes used in each of the pilot studies.

### **Uncertainty representation**

In general, evaluations on many attributes (particularly those relating to product or service experience) may be expected to be uncertain. They may fluctuate from trial-to-trial,

Dataset 1 Toothpaste	Dataset 2 Laundry detergent
High quality	High quality
Trusted	Trusted
Value for money	Value for money
Becoming more popular	Becoming more popular
Positive word-of-mouth	Positive word-of-mouth
Prominent packaging	Prominent packaging
Pleasant taste	Gets clothes clean
Keeps breath fresh	Quite expensive
Cleans teeth	Nice fragrance
Quite expensive	Gentle on clothes
Protects teeth	

Table 5.4: Lists of attributes used to evaluate alternatives in the two market research studies

either because of differences in the product or service over time, or because of environmental differences which affect product or service performance. Respondents are also allowed to evaluate alternatives which they are aware of but have not used, in which case some uncertainty is bound to be present. However, constraints on the survey process mean that it is generally not feasible to collect detailed information regarding uncertainty. Market research surveys are typically long (often running to 100 or more questions; only small parts of the full surveys are used here) and require substantial time and effort from respondents, who are generally not compensated. They are also not facilitated in any meaningful way. In the vast majority of studies (including the two discussed here), respondents evaluate alternatives on attributes by giving an *expected* rating on some (usually binary or Likert) constructed scale. In this case, the choice of uncertainty format is essentially constrained by difficulties of data collection.

Because no attribute importance information is gathered, the SMAA approach is an obvious candidate as a ‘decision’ model. The lack of weight information is due mostly to a tradition which distrusts collecting explicitly “stated” importance weights, and the difficulty in reliably collecting weight information using unfacilitated surveys. In the absence of weight information, the acceptability index  $b_i^r$  gives an indication of the relative number of possible weight vectors that support the selection of  $a_i$  in rank  $r$ . In its use as a metric for brand performance, it describes a model in which an individual’s preferences (the relative weights allocated to each attribute) will show some variation, as a function of (a) internal changes in what is desired, (b) changes in purchase-to-purchase contextual factors, and (c) other forms of essentially random variation. The source of the variation is not important for the purposes of the analysis. If the variation is entirely random, then the acceptability indices represent the long-run purchase probability for each brand.

Ehrenberg and his co-authors have shown in a series of books and papers [93, 106, 259] that purchase behaviour can be fruitfully modelled as occurring “in an *as if* random manner” at the individual level even if true randomness is only a feature of the model. This motivates the use of the acceptability index as a predictive measure of expected relative purchase frequencies, both at the individual and aggregate level.

### Designing the validation study

In both studies, respondents completed an initial survey at  $t_0$  in which they identified:

- $B_1(t_0)$ : the set containing those brands the respondent was aware of
- $B_2(t_0)$ : the set containing those brands the respondent had purchased in the last 3 months
- $B_3(t_0)$ : the set containing those brands the respondent would consider buying, including those brands in  $B_2(t_0)$
- Sat: overall satisfaction with each brand in  $B_1$ , evaluated on a 10-point scale from 1 (“terrible”) to 10 (“perfect”)
- PI: the intention to purchase each of the brands in  $B_3$  within the next few weeks, evaluated on a 5-point scale from 0 (“definitely will not buy”) to 1 (“definitely will buy”)
- AdR: a binary indicator of whether or not the respondent could remember seeing any advertising for each of the brands in  $B_3(t_0)$
- SoP( $t_0$ ): the relative proportion of the respondent’s last 10 purchases allocated to each of the brands in  $B_2(t_0)$

Respondents were also shown a list of relevant attributes and were asked to associate brands from  $B_1(t_0)$  with each of the attributes. Respondents were explicitly told that they could select one, more than one, or no brand at all for each attribute statement. There were 11 and 10 attributes in the studies of the UK toothpaste and Spanish laundry detergent markets respectively. These responses constitute the expected attribute evaluations ( $z_{ij} = 1$  if brand  $a_i$  was selected as possessing attribute  $c_j$  and  $z_{ij} = 0$  otherwise). In both datasets, one of the attributes was negatively worded (“quite expensive”). Responses on this attribute were simply coded to take on the value  $-1$  if a brand possessed the negative attribute and zero otherwise i.e. preferences are increasing in all attributes. Different acceptability indices  $b_i^r$  for  $r = \{1, 3, 5\}$  were calculated on the basis of 20000 random weight vectors generated for each respondent, in line with results in [252]. Weights are generated so as to be uniformly distributed and sum to one.

In addition, a further measure of preference for each brand was constructed from an equally-weighted linear combination of the attribute values i.e.  $\sum_j z_{ij}$ . This modelled is labelled ‘Dawes’ after previous research by that author showing that this model can be an excellent and robust representation of preferences [74]. By setting all weights equal to one, the measure can be thought of as representing the net number of positive attributes associated with each brand. It is also consistent with the other measures in that it can be constructed from a simple set of survey questions without reference to actual or claimed purchase frequencies.

A follow-up survey of all respondents at  $t_1$ , approximately 6 months after the initial survey, was used to gather information on the set of brands the respondent had purchased in the 3 months prior to the  $t_1$  survey i.e.  $B_2(t_1)$ , as well as the relative proportion of the respondent’s previous 10 purchases attributed to each of the brands in  $B_2(t_1)$  i.e.  $\text{SoP}(t_1)$ . For ease of presentation, the analysis section to follow focuses on three target brands in each dataset – the brand with the highest penetration, a ‘medium-sized’ brand selected at random from those brands with penetrations from 20%–50%, and a ‘small’ brand selected at random from those brands with penetrations from 5%–10%. Zero and 11 dropouts were reported in the UK toothpaste and Spanish laundry detergent surveys respectively, leaving effective sample sizes of  $n_1 = 281$  and  $n_2 = 350$ . In addition to the relative purchase frequencies described above, the initial and follow-up survey information was used to determine which respondents had defected from brands used at  $t_0$  by the time of the follow-up survey at  $t_1$ .

### 5.6.3 Results

#### Associations between the acceptability index and relative purchase frequency

Tables 5.5 and 5.6 show the relationship between predictions of aggregate purchase frequency i.e. a proxy for market share, obtained using five predictive measures and actual market share at  $t_0$  and  $t_1$ , for all brands with market shares greater than 5%, in each market. Aggregated brand-level measures are obtained by averaging each measure over all respondents. For brevity only the best-performing acceptability index  $b_i^1$  is shown. All measures make use of the relative rather than absolute performance levels i.e. relative satisfaction, advertising recall, net number of positive attributes, and purchase intentions have been used. This simply reflects a preference model in which purchase probabilities are proportional to absolute levels of satisfaction, advertising recall, net number of positive attributes, or purchase intentions respectively, and is necessary for market shares to sum to 100% at both an individual and aggregate level. The acceptability index is an inherently relative quantity and so does not require any standardisation. Finally,  $\bar{\delta}(t_k) = \sum_{i=1}^n |\Psi_i - \text{SoP}_i(t_k)|/n$  denotes the average absolute deviation between actual

market share  $\text{SoP}_i(t_k)$  and predicted market share  $\Psi_i$  over all brands in each market. Significant differences between these values and the values of  $\bar{\delta}(t_k)$  obtained using the acceptability index are flagged using a single, double, or triple asterisk superscript to denote significance at the 10%, 5% or 1% level respectively (the nonparametric Wilcoxin matched pairs test was used for this purpose since distributions of market shares are strongly skewed to the right).

Rank	SoP( $t_0$ )	SoP( $t_1$ )	Sat	AdR	Dawes	PI	$b_i^1$
1	44.8%	52.4%	28.7%	16.9%	30.3%	34.1%	46.2%
2	17.0%	13.7%	16.8%	14.1%	14.9%	16.1%	17.9%
3	9.9%	9.4%	9.9%	19.8%	9.3%	10.3%	7.7%
4	7.0%	8.6%	4.4%	2.6%	3.6%	5.4%	4.5%
5	5.9%	6.5%	11.5%	4.2%	9.0%	9.3%	5.8%
$\bar{\delta}(t_0)$			2.3%	4.4%***	2.4%***	1.6%	0.8%
$\bar{\delta}(t_1)$			3.3%	5.2%***	3.2%**	2.5%	1.5%

Table 5.5: Associations between survey measures and overall market share for all brands with market share greater than 5% in Dataset 1

Rank	SoP( $t_0$ )	SoP( $t_1$ )	Sat	AdR	Dawes	PI	$A_i^{(1)}$
1	30.1%	27.6%	25.1%	21.1%	26.7%	24.7%	30.9%
2	19.2%	21.5%	11.7%	1.8%	13.7%	16.1%	15.3%
3	11.8%	11.0%	14.2%	9.7%	12.8%	13.6%	11.1%
4	10.0%	9.2%	13.9%	11.7%	12.9%	11.6%	10.6%
5	6.3%	7.7%	7.9%	10.2%	7.7%	7.3%	7.1%
6	6.0%	6.7%	5.3%	4.8%	5.6%	5.5%	5.6%
$\bar{\delta}(t_0)$			2.4%**	5.4%***	1.7%	1.6%	1.1%
$\bar{\delta}(t_1)$			2.5%*	5.3%***	1.8%	1.8%	1.5%

Table 5.6: Associations between survey measures and overall market share for all brands with market share greater than 5% in Dataset 2

The first-rank SMAA acceptability index offers very accurate predictions of aggregate relative purchase frequencies, particularly at  $t_0$  but also at  $t_1$ . The average deviation between the market share predicted by the acceptability index and the actual market share is smaller than any of the other measures, although the difference is only significant for advertising recall and the relative net number of positive attributes (Dataset 1) and satisfaction and advertising recall (Dataset 2). The former result is important because it shows that the acceptability index is demonstrably superior to the simpler-to-implement net count of the number of positive attributes. It is also noteworthy that in Dataset 1 only the acceptability index is able to accurately predict the market share of the leading brand for this market. The results in Table 5.5 and 5.6 provide a strong indication of the ability of the acceptability index to model buying behaviour at the aggregate level, particularly

in relation to other existing measures.

Having considered purchase behaviour in the aggregate, one can turn to the more difficult task of modelling individual buying behaviour. Tables 5.7 and 5.8 show correlations between each of the seven measures (three SMAA acceptability indices are now used, plus the four other measures) and the shares of purchases allocated to each brand at  $t_0$  and  $t_1$ , for each market. Again, all measures use the relative rather than absolute performance levels. Significance at the 5% and 0.5% level is denoted using a single and double asterisk superscript respectively.

Outcome	Brand	Sat	AdR	Dawes	PI	$A_i^{(1)}$	$A_i^{(3)}$	$A_i^{(5)}$
SoP( $t_0$ )	Large	0.65**	0.08	0.62**	0.85**	0.76**	0.61**	0.57**
	Medium	0.78**	0.48**	0.76**	0.88**	0.69**	0.58**	0.50**
	Small	0.61**	0.19**	0.67**	0.75**	0.68**	0.49**	0.45**
SoP( $t_1$ )	Large	0.34**	0.03	0.42**	0.49**	0.52**	0.44**	0.39**
	Medium	0.63**	0.04	0.69**	0.72**	0.68**	0.47**	0.40**
	Small	0.15*	0.16*	0.16*	0.25**	0.34**	0.43**	0.43**

Table 5.7: Associations between survey measures and share of purchases in Dataset 1

Outcome	Brand	Sat	AdR	Dawes	PI	$A_i^{(1)}$	$A_i^{(3)}$	$A_i^{(5)}$
SoP( $t_0$ )	Large	0.62**	0.24**	0.69**	0.80**	0.67**	0.58**	0.55**
	Medium	0.65**	0.20**	0.76**	0.83**	0.73**	0.66**	0.65**
	Small	0.70**	0.18**	0.64**	0.75**	0.60**	0.58**	0.56**
SoP( $t_1$ )	Large	0.34**	0.11*	0.43**	0.46**	0.43**	0.31**	0.30**
	Medium	0.33**	0.12*	0.50**	0.44**	0.49**	0.38**	0.37**
	Small	0.30**	0.09	0.34**	0.37**	0.41**	0.29**	0.28**

Table 5.8: Associations between survey measures and share of purchases in Dataset 2

All acceptability indices are significantly related to relative purchase frequency at the 0.5% level at both  $t_0$  and  $t_1$ , providing strong support in favour of the first research hypothesis. With the exception of the small brand in Dataset 1 at  $t_1$ , the strength of association between purchase frequencies and the acceptability indices decreases as more of the rank order is considered. In Dataset 1, the first-rank acceptability index shows a higher correlation with relative purchase frequency at  $t_0$  than either the net number of positive attributes or overall satisfaction for two of three brands, although all are markedly lower than the correlations obtained using purchase intentions. At  $t_1$ , correlations involving  $b_i^1$  are again higher than the net number of positive attributes and overall satisfaction, this time for all three brands and to a greater extent than at  $t_0$ . The first-rank acceptability index exhibits higher correlations with relative purchase frequencies at  $t_1$  than purchase intentions do for two of the three brands. The results obtained from the second dataset

largely echo those obtained from the first. All acceptability indices are highly significantly related to relative purchase frequency at both  $t_0$  and  $t_1$ , with the strength of associations declining as the number of ranks considered increases from one to three to five. Purchase intentions are most strongly correlated with relative purchase frequency at  $t_0$  but not at  $t_1$ , where purchase intentions, the relative net number of positive attributes, and the first-rank acceptability index are all roughly equally-strongly correlated with purchase frequency.

### Associations between the acceptability index and defection behaviour

Tables 5.9 and 5.10 show the relationship between each of the seven measures and the likelihood of defection, for each market respectively. In these tables, the absolute rather than relative levels of satisfaction, advertising recall, and net number of positive attributes have been used. Within each brand, the first two rows give the average value of each measure among non-defecting and defecting users respectively. The third row gives a  $Z$ -statistic obtained from a non-parametric Mann-Whitney test for equality of means. Results that are significant at the 5% level are superscripted with a single asterisk and results that significant at the 0.5% level are superscripted with a double asterisk. The fourth row shows the odds ratio (OR) associated with a logistic regression of defection on each measure. The fifth and final row in each sub-table shows the  $Z$ -statistic associated with each odds ratio, and is superscripted to denote statistical significance in the same way as described above.

		n	Sat	AdR	Dawes	PI	$A_i^{(1)}$	$A_i^{(3)}$	$A_i^{(5)}$
Large brand	$\bar{X}_{\text{Stay}}$	181	8.28	0.33	7.83	0.82	0.65	0.81	0.84
	$\bar{X}_{\text{Defect}}$	16	7.81	0.13	3.63	0.56	0.18	0.36	0.43
	$Z_{\bar{X}}$		1.87	1.70	4.10**	3.80**	3.88**	3.95**	4.23**
	OR		0.73	0.29	0.71	0.03	0.05	0.08	0.10
	$Z_{\text{OR}}$		-1.65	-1.61	-4.13**	-3.59**	-3.59**	-3.99**	-3.85**
Medium brand	$\bar{X}_{\text{Stay}}$	36	8.58	0.50	6.50	0.83	0.42	0.67	0.71
	$\bar{X}_{\text{Defect}}$	23	8.26	0.39	4.48	0.65	0.11	0.39	0.47
	$Z_{\bar{X}}$		0.48	0.81	2.11*	2.74*	2.51*	2.28*	1.96*
	OR		0.61	0.64	0.84	0.08	0.04	0.24	0.32
	$Z_{\text{OR}}$		-1.42	-0.82	-2.08*	-2.36*	-2.72*	-2.30*	-1.92
Small brand	$\bar{X}_{\text{Stay}}$	14	7.79	0.14	5.36	0.73	0.18	0.39	0.47
	$\bar{X}_{\text{Defect}}$	15	7.00	0.07	3.87	0.47	0.08	0.24	0.30
	$Z_{\bar{X}}$		1.24	0.66	1.19	2.61*	1.22	0.92	0.95
	OR		0.69	0.43	0.89	0.02	0.30	0.40	0.41
	$Z_{\text{OR}}$		-1.31	-0.66	-1.08	-2.18*	-0.87	-1.00	-1.02

Table 5.9: Associations between survey measures and defection behaviour in Dataset 1

The second research hypothesis, stating that probabilities of defection should be inversely related to the SMAA acceptability indices, is supported for two of the three brands in



		n	Sat	AdR	Dawes	PI	$A_i^{(1)}$	$A_i^{(3)}$	$A_i^{(5)}$
Large brand	$\bar{X}_{\text{Stay}}$	116	7.89	0.63	5.40	0.79	0.54	0.82	0.88
	$\bar{X}_{\text{Defect}}$	39	4.79	0.46	2.95	0.30	0.17	0.52	0.55
	$Z_{\bar{X}}$		4.70**	1.83	4.31**	6.60**	5.40**	2.87**	4.30**
	OR		0.63	0.5	0.74	0.01	0.04	0.17	0.14
	$Z_{\text{OR}}$		-4.94**	-1.83	-4.09**	-6.17**	-4.59**	-3.93**	-4.34**
Medium brand	$\bar{X}_{\text{Stay}}$	44	7.57	0.45	4.61	0.73	0.44	0.72	0.79
	$\bar{X}_{\text{Defect}}$	34	3.26	0.50	1.56	0.20	0.07	0.23	0.29
	$Z_{\bar{X}}$		5.15**	-0.40	4.69**	5.92**	4.94**	4.64**	4.18**
	OR		0.56	1.2	0.63	0.01	0.02	0.07	0.08
	$Z_{\text{OR}}$		-4.05**	0.40	-3.81**	-4.85**	-3.43**	-4.36**	-4.32**
Small brand	$\bar{X}_{\text{Stay}}$	13	7.23	0.69	3.85	0.79	0.28	0.63	0.70
	$\bar{X}_{\text{Defect}}$	14	1.71	0.57	1.29	0.13	0.12	0.19	0.25
	$Z_{\bar{X}}$		3.26**	0.64	3.04**	4.23**	2.29*	2.47*	2.14*
	OR		0.56	0.59	0.70	0.001	0.19	0.07	0.10
	$Z_{\text{OR}}$		-2.96**	-0.65	-2.08*	-2.92**	-1.23	-2.44*	-2.38*

Table 5.10: Associations between survey measures and defection behaviour in Dataset 2

Dataset 1 and all three brands in Dataset 2 (although with some discrepancies in which index is most significant). All relationships are in the hypothesised direction – average acceptability indices for all brands are smaller among defecting users than among non-defecting ones, and odds ratios are all less than one, indicating a reduced risk of defection as the acceptability index increases. For example, the acceptability index  $b_i^1$  for the largest brand in the market in Dataset 1 indicates that on average 65% of all preferences supported the brand as ‘best’ among those who did not defect. In comparison just 18% of all preferences supported the brand as best among those who defected within the next six months. Similar results are obtained for the medium-sized brand. Odds ratios indicate that those with acceptability indices around 0 (indicating a dominated alternative) are around 20 times more likely to defect than those with acceptability indices around 1 (indicating a dominating alternative).

The SMAA acceptability index compares favourably with the other attitudinal equity measures. In Dataset 1 overall satisfaction is non-significant for all three brands, the Dawes measure is of roughly equal significance with the first-ranked acceptability index, and purchase intentions are the only other measure showing a consistently significant relationship with defection behaviour. In Dataset 2 overall satisfaction is significant for all three brands, the Dawes measure is again of roughly equal significance with  $b_i^1$ , and purchase intentions are strongly associated with future defection behaviour.

### Associations between the acceptability index and changes in relative purchase frequency

Tables 5.11 and 5.12 show the partial correlations between the seven measures with relative purchase frequencies at  $t_1$ , after controlling for the effect of purchase share at  $t_0$ . It is also possible to test hypothesis 3 by regressing changes in purchase frequency i.e.  $SOP(t_1) - SOP(t_0)$ , on both relative purchase frequency at  $t_0$  and each of the predictive measures. The conclusions to be drawn are the same using either analysis but the partial correlations more concisely address the question of whether an association exists between a measure and changes in relative purchase frequencies over time. As before, significance at the 5% and 0.5% level is denoted using a single and double asterisk superscript respectively.

Outcome	Brand	Sat	AdR	Dawes	PI	$A_i^{(1)}$	$A_i^{(3)}$	$A_i^{(5)}$
SoP( $t_1 t_0$ )	Large	-0.06	-0.02	0.09	0.01	0.17**	0.15*	0.15*
	Medium	-0.02	-0.21**	0.18**	-0.01	0.29**	0.00	0.00
	Small	-0.16*	0.01	-0.19**	-0.13*	0.06	0.27**	0.29**

Table 5.11: Associations between survey measures and share of purchases in Dataset 1

Outcome	Brand	Sat	AdR	Dawes	PI	$A_i^{(1)}$	$A_i^{(3)}$	$A_i^{(5)}$
SoP( $t_1 t_0$ )	Large	-0.04	-0.03	0.05	-0.02	0.06	-0.04	-0.04
	Medium	-0.04	0.01	0.15**	-0.04	0.15**	0.02	0.02
	Small	-0.06	0.01	0.04	0.02	0.17**	0.01	0.00

Table 5.12: Associations between survey measures and share of purchases in Dataset 2

The first-rank acceptability index is the only measure that exhibits reasonably consistent significant correlations with changes in relative purchase frequencies over time. Although correlations are not high, the correlations obtained with  $b_i^1$  are positive and significantly different from zero at the 0.5% level for two of the three brands in both datasets. Correlations for the two other acceptability indices are also significant in Dataset 1 for two of the three brands. These correlations provide some support in favour of the third research hypothesis, to the effect that changes in the relative purchase frequency are positively related with the SMAA acceptability index. No other measure approaches a similar level of consistent statistical significance. In Dataset 1, the net number of positive attributes is highly significant for the medium-size brand and is significant but in the wrong direction for the small brand. The correlations obtained using purchase intentions are either not significant or in the wrong direction. In Dataset 2, the relative net number of positive attributes exhibits significant correlations for the medium-sized brand only.

#### 5.6.4 Follow-up

As a result of the two validation studies, the stakeholders in the market research company were satisfied that a first-rank SMAA acceptability index using expected values was able to give a measure of brand performance that correlated well with observed purchase behaviour at the individual and aggregate (brand) level. The index has since been used in more than 10 further studies for several large client brands in South Africa, with representatives from a number of product and service categories: consumer goods (McDonald's, KFC), retail (Pick 'n Pay), travel (South African Airways), and technology (MWeb).

### 5.7 Conclusions

It was clearly possible to employ various simplified uncertainty formats in practical decision problems. In the first application, the inclusion of quantiles around the original estimates of expected values was felt to add significant value to the analysis. Two alternatives that were dominated at their mean values (and hence had no central weight vectors) became relatively more attractive once quantiles were included. This provided a richer set of electricity user "typologies". The group also felt it was useful to be able to assess the extent to which the attractiveness of options depend on the manner in which they are applied. Media for options like the hotbox which are much more attractive at the best-case quantile (which implies a 'correct' application, among other factors beyond the control of the group) should be accompanied by greater educational material than those options which give relatively similar performance over quantiles, for example.

In the second application it was shown that, although explicit risk measures like the variance and probability of poor performance were found in earlier chapters to lead to poorer decisions than other uncertainty formats, in some application areas these measures are the way in which uncertainty is most naturally expressed. This makes the point that although the previous chapters suggest that facilitators should hold certain preconceptions about uncertainty formats – particularly that quantiles are a good 'default' option – they should certainly be flexible enough to accommodate whatever types of uncertainty representation seem most appropriate for a given decision maker and problem context. In the third application, the choice of uncertainty format was essentially constrained to be expected values by virtue of the data collection process. The use of expected values to represent uncertain attribute evaluations is perhaps more common than is apparent from the MCDA literature; even if not done explicitly, where uncertainty is perceived to be limited in magnitude or resources are limited, there is surely a tendency to use expected values. In this particular application, decision makers were satisfied enough with the use of expected values to continue to use them in similar studies in the future.

In terms of difficulties, the assessment of quantiles in the first application did take some time and was experienced as moderately difficult (although also useful as a learning process). Although 5% and 95% quantiles were requested, the researcher did not feel that he was able to deliver estimates to this level of precision, and preferred to think of evaluations as “almost” worst- and best-case performances respectively. The interpretation of quantile weights required some care, and the joint interpretation of attribute and quantile weights i.e. of the  $w_{jk}$ , was experienced as difficult. Instead, the decision maker found it easier to interpret the attribute importance weights for each quantile ( $w_{j|k}$ ), and the quantile weights ( $w_{q_k}$ ), separately. Although no elicitation of weight information was performed, the same difficulties and responses might be expected to apply there too. The evaluations on the explicit risk attributes in the second application were assessed via a computer simulation model and no subjective information was required from the client. Central weight vectors in the second application were therefore relatively straightforward – in that they show only two (explicit risk) attributes – and were easily digested and found to convey useful information.

One of the unintended features of the three applications is the use of a SMAA model. The SMAA models address *internal* uncertainty, particularly uncertainty relating to the preferences of the decision maker, by presenting information about the consequences of particular preferences rather than by eliciting those preferences directly (the approach taken by most MCDA methods). It may be that many of the conditions that suggest the use of the SMAA approach for resolving internal uncertainty might also suggest the use of a simplified format for resolving external uncertainty. In particular, all three of the applications reported in this chapter were characterised by constraints on the time and effort that stakeholders were willing to devote to the decision process. In such cases the simplified SMAA models developed in Section 5.3 may be useful. Decision makers in all applications found the acceptability indices (within quantiles in the first application, overall in all applications) intuitive, easy to interpret, and useful in terms of the information conveyed.

# Chapter 6

## Conclusions

In this final chapter I collect together the conclusions drawn in the previous three chapters and state these as a number of findings and related recommendations for the support of decisions involving uncertain attribute evaluations. The emphasis in doing so is on providing insights that are practical and general enough to be useful to facilitators and other practitioners of decision analysis. Key findings are listed in bold below and elaborated on in the text. The sources on which each finding is based (simulation experiment (Chapter 3), choice experiment (Chapter 4), or real-world application (Chapter 5); denoted **S**, **C** and **A** respectively) are given in square brackets following the finding. Despite the different methodological approaches used in the three chapters, where it has been possible to base a finding on multiple approaches there is substantive agreement between the approaches. Following the findings and recommendations, some questions for future research are listed.

### 6.1 Findings and recommendations

**Finding 1a: The use of any of the simplified uncertainty formats considered in this thesis (rather than probability distributions) can be justified. [S,C,A]**

**Finding 1b: Placing elements of simplified uncertainty formats (scenarios, quantiles, variances and other explicit risk attributes) in the second level of the objectives hierarchy can offer a useful uncertainty-orientated view of a decision problem. [A]**

Perhaps the most important message emerging from the simulation and choice experiment results is that, given the time and effort required to implement a MAUT model, the use of several simplified uncertainty formats can be justified from an accuracy-effort perspective. Simulation results gave average utility losses for error-free simplified models of between 0.004 for quantiles and 0.03 for scenarios. In the choice experiment, even though choices were artificial and made without facilitation the average utility loss ranged between 0.12 for quantiles and 0.26 for expected values. Subjects in the choice experiment rated all sim-

plified uncertainty formats except variances easier to use than probability distributions. They also selected poorer alternatives using probability distributions than using any simplified uncertainty format except expected values, suggesting real difficulty in correctly employing probability information. Helping decision makers address this difficulty is part of the responsibility of decision analysis, and a process of facilitation through the steps of assessing, validating and using probabilities in a MAUT (or similar) model is almost certainly the best way to do this. But sometimes it will simply not be possible to carry out this process well. The applications in Chapter 5 illustrate decision problems where the aim is to narrow down the set of potential alternatives to a smaller shortlist and decision makers are unwilling to expend much time and effort in doing so, but these are by no means the only possible reasons for preferring to use a simplified model. Results suggest that it is better to correctly apply any of the simplified decision models than to apply a probability-based model poorly.

This still leaves practitioners with considerable scope to use whatever model they feel is least likely to be applied poorly. Other results (discussed in the next finding) suggest that the largest determinant of decision accuracy is the extent to which assessment errors are avoided. Practitioners should therefore primarily consider the knowledge and experience of the decision makers they are facilitating, and choose a simplified uncertainty format that they and the decision makers are comfortable working with.

Although not so much a finding as an observation based on past literature, considering uncertainty elements (scenarios, quantiles, variances or other explicit risk attributes) in the second level of the objectives hierarchy (rather than at the lowest level) imposes a quite different but potentially useful view of uncertainty on MCDA. Uncertainty is not viewed as something which must simply be aggregated over to provide a valid indicator of performance, but as something that can be decomposed into elements in much the same way that overall performance can be decomposed into attributes. In this view, assessments of alternatives' performances in each uncertainty element provide useful information in their own right. This has an established history in scenario planning, but was also found to be useful in the quantile models used in the electricity saving application reported in Chapter 5.

**Finding 2a: Deviations from linearity in the marginal utility functions increase the attractiveness of MAUT relative to all of the simplified models. [S]**

**Finding 2b: Increases in the size of a decision problem (number of alternatives and attributes) increase the attractiveness of all of the simplified models relative to MAUT. [S]**

In various parts of this thesis the choice of whether to use a simplified model (and if so, which model to use) has been viewed as a meta-choice problem in which decision quality (accuracy), effort, and contribution to “learning” are the primary attributes. Some characteristics of a decision problem may therefore serve to make certain decision models more or less attractive relative to others by differentially affecting any of the three attributes. The simulation results indicate that the approximation of MAUT provided by all the simplified models deteriorates as marginal utility functions deviate from linearity. If it can be assumed that the shape of the utility functions does not affect the *relative* amount of effort required (or learning imparted) by any of the models, which seems reasonable, then deviations from linearity act to increase the relative attractiveness of MAUT in the meta-choice problem (note that this does not imply an absolute preference for MAUT over any of the simplified models).

Similarly, the simulation results indicate that approximation accuracy (in the form of utility loss) is almost unaffected by the number of alternatives and attributes present. If it can be assumed that effort increases with problem size and that the MAUT model is more sensitive to increases in problem size than any of the simplified models, then a greater number of alternatives or attributes acts to increase the relative attractiveness of all simplified models in the meta-choice problem (assuming relative amounts of “learning” are unaffected by problem size). Increases in problem size will also favour those simplified models requiring less effort *per alternative* or attribute than others, but not enough is known to draw any firm conclusions here.

**Finding 3a: Errors in the assessment of inputs to all simplified models cause substantial deteriorations in model accuracy. [S]**

**Finding 3b: Quantile and scenario models are slightly more robust to assessment errors than expected value and variance models. [S]**

The simulation results indicate that avoiding errors when assessing an uncertainty format is more important to decision accuracy than any of the other experimental factors (for the ranges of parameter values used), including which uncertainty format is used. Practitioners must ensure that any uncertainty information is assessed as accurately as possible. This is particularly true of models using expected values and explicit risk attributes, which (under the assumption of random assessment errors) are more sensitive to assessment errors than the scenario and quantiles models, which use a greater number of inputs to measure risk and thus benefit to a greater extent from “error cancellation” [151].

Practitioners should therefore consider the likelihood of assessment errors when deciding which uncertainty format to use. Previous literature is in basic agreement that errors in

the assessment of a particular format (a) are less likely if the decision maker has prior knowledge and experience of that format, (b) are more likely with variances or probabilities than with expected values and quantiles (little is known about scenarios), and (c) are more likely with more extreme quantiles. Some time ago Amos Tversky wrote of a need to develop decision aids “more responsive to the complexities and limitations of the human mind” [256]. The fact that errors are less likely using quantiles than variances *and* that quantiles are more robust to errors than variances surely constitutes evidence in favour of the former and against the latter. This is discussed further in the next two findings.

**Finding 4: A model using quantiles is the most promising of the simplified models and should be used as a ‘default’ option in the absence of other information. [S,C,A]**

Although one of the more important messages emerging from the results is that it is not possible (because of circumstantial factors like decision maker experience which affect ease of use and assessment error) to say that one uncertainty format is definitively “better” than another, this does not imply that all uncertainty formats should be equally favoured *a priori* by practitioners. The advocacy of quantiles is based on three results obtained in this thesis. Firstly, the simulation results show that a quantile model gives the best approximation to MAUT. Secondly, the choice experiment results show that the same holds in unfacilitated decision making, and also that quantiles were judged relatively easier to use than probability distributions. Thirdly there is the fact that it was possible to apply a SMAA model using quantiles and generate a successful outcome without any significant problems. In addition, previous research has shown that quantiles can be assessed relatively accurately by decision makers [102], and they are already a standard tool in decision analysis (in the assessment of probability distributions e.g. [238]).

The results indicate that three quantiles are sufficient to obtain good results. The quantiles included (in addition to the median, which appears standard) should not be as extreme as minima and maxima, and the 5% and 95% quantiles were used with good results. Although decision makers find the minima and maxima easier to work with they tend to lead to lower accuracy, and the presence of the minima in particular can increase risk aversion dramatically, as seen in the choice experiment. Some insight may be gained, as was the case for the electricity saving application presented in Chapter 5, by viewing quantiles as higher-level criteria situated in the second-tier of the objectives hierarchy (see Figure 2.2) and allowing decision makers to evaluate and compare alternatives at their ‘worst-case’, ‘best-case’ and some ‘intermediate’ levels of performance. In aggregating performances over quantiles, the simulation results showed that fidelity to MAUT is best served by using the weights originally proposed by Keefer and Bodily in [144]: the median receives a



weight of 0.63 and the remainder is shared between the extreme quantiles. It is recommended that these weights be used as ‘default’ values to be replaced only where it is clear that the decision maker values performance in the quantiles differently.

**Finding 5: A model using variances is the least promising of the simplified models and should not be used unless there are compelling reasons to do so. [S,C]**

Again with the caveat that in a particular decision problem there may be circumstantial reasons for favouring the use of variances, several of my (and others’) results suggest that variances are a relatively poor option for representing uncertainty in decision analysis. The simulation results showed that a model based on expected values and variances performed no better than a model based on expected values alone, and was less robust to assessment error than all models except the expected value model. The choice experiment also showed that in multi-criteria decisions variances are both less accurate and more difficult to use than quantiles, fuzzy numbers, and scenarios. Previous research indicates that variances are difficult to interpret and assess numerically [102], and trade-offs involving variances would appear considerably more difficult to reason about than trade-offs involving risk measured on the same scale as the underlying attribute (e.g. quantiles, scenarios). The latter difficulty can be avoided if variances are weighted using Kirkwood’s weights [150]. The simulation results indicated that a variance model using Kirkwood weights performed much better than a variance model in which variance weights were a constant multiple of attribute importance weights, but only very marginally better than an expected value model. Taken together this suggests that variances should only be used (a) in conjunction with Kirkwood’s weights, (b) where there are compelling reasons to use variances to represent uncertainty, and (c) where the conditions favouring the variance model over an expected value model (approximately normally distributed attribute evaluations and utility functions not deviating too much from linear [150]) are satisfied.

**Finding 6: Scenarios result in relatively poor approximations of MAUT and are primarily devices for helping decision makers to gain a better understanding of the causes of uncertainty and their consequences. [S,C]**

The simulation results indicate that scenario-based models only give approximations of MAUT that are comparable with models using other uncertainty formats under quite specific requirements for the construction of scenarios – a large proportion of the possible futures must be captured; and in each scenario the average performance of each alternative in that scenario must be assessed. These requirements, which ensure that expectations (over scenarios) approximate expected attribute values, were fulfilled by construction in the choice experiment, but fulfilling them generally implies a process quite different to the

philosophy of scenario planning, which often advocates constructing a small number of extreme scenarios without regard for any implications for later aggregation.

This leaves two options for scenario-based MCDA. Either scenario construction can ignore established conventions and proceed along the lines just described, or scenario planning principles can be integrated with MCDA while acknowledging that approximating expected utility is not of great importance for the decision process. The latter is recommended on the basis that it seems far more likely to be accepted by most decision makers. Scenario planning is already established as a popular method in areas of strategic decision making where MCDA could possibly be applied. The onus would probably be viewed as being on MCDA to integrate scenario planning principles rather than the other way around, and failing to do so may alienate many users of scenario planning from MCDA. Increasing the importance of “learning” about uncertainty at the expense of the importance accorded to approximating expected utility is unlikely to be viewed as untenable by users of scenario planning, which was originally developed by changing the goal of forecasting from ‘approximating expected value’ to ‘learning and contingency planning’.

Once scenarios have been constructed, the choice experiment as well as applications in the literature [197] indicate that decision makers are able to work and reason with these, so that decision maker learning about preferences is unlikely to suffer from the use of scenarios. The main expected benefit then, is a greater understanding of the underlying causes and consequences of uncertainty (this is well-established in the literature [260], but has not been tested in this thesis); it is the importance accorded to this benefit which should largely determine whether scenarios are used.

## 6.2 Questions for future research

Much work has been done on how people assess and reason about uncertain outcomes, and on how they assess and reason about conflicting objectives, but much less is known about how these two elements interact with one another. This thesis has attempted largely exploratory work on this interaction, mainly from the perspective of controlled experimental tests rather than real-world case studies. This leaves much for future research. Some of those questions that appear more important or immediate are listed below. Methodologies which might prove useful in investigating each question are given in square brackets following each question (again using **S**, **C** and **A** to denote simulation experiments, choice experiments, and real-world applications respectively).

**Question 1a: Which uncertainty formats do decision makers find easier to use in real-world prescriptive decision problems? [A]**

**Question 1b: Which uncertainty formats do decision makers perceive as helping them to make better decisions in real-world prescriptive decision problems?** [A]

Assessing individual case studies, even in an action-research mode, brings its own complications, but an important task is to systematically evaluate the simplified models in ‘traditional’ real-world MCDA problems. This would allow one to (a) assess the ease with which decision makers make trade-off assessments when different uncertainty formats are used, (b) possibly disconfirm the observations drawn from controlled environments in this thesis, (c) assess the extent to which using more concise summaries of uncertainty than probability distributions affects decision makers’ belief in the process.

**Question 2: Does uncertainty format and the number of objectives used interact to affect decision accuracy if more objectives are used?** [C,A]

The amount of information made available to the decision maker (and hence uncertainty formats) affects accuracy both by placing different cognitive ‘loads’ on decision makers and by obscuring important information (in the case of the simplified uncertainty formats). The choice experiment in Chapter 4 indicated no interaction effect between uncertainty format and the number of objectives used, suggesting that whatever the difference between the ‘load’ and ‘information loss’ for a particular model, these differences are constant over objectives *for all models*. If this holds for any number of objectives it is a important result, since it implies that no model becomes relatively more or less favoured (from an accuracy perspective) as the number of objectives increases.

**Question 3a: How often do preferences (utility functions, importance weights) differ over uncertainty elements (particularly scenarios and quantiles)?** [A]

**Question 3b: Where preferences do differ over uncertainty elements, is it better to assess alternative-uncertainty element ‘meta-alternatives’ over attributes or to assess alternatives over attribute-uncertainty element ‘meta-attributes’?** [A]

One of the more intriguing possibilities arising from the use of simplified models is the possibility of allowing preferences to vary over elements of uncertainty (scenarios, quantiles, means and variances or other explicit risk attributes). There is now some evidence (in [197] and the first application of Chapter 5) that this may be of benefit in some decision problems. It remains an open question how often preferences differ over uncertainty elements and how preferences should be elicited where they do differ. Montibeller [197] has also called for “experimental evaluations of the cognitive burden in eliciting preferences in

both approaches”.

**Question 4: Do the same results found in this thesis extend to other schools of MCDA (particularly outranking and reference point methods)? [S,C,A]**

All models and results in this thesis are based upon value function methods. It is possible that at least some of the conclusions drawn might differ if other schools of MCDA are used. Although because of the lack of an axiomatic foundation in other schools of MCDA, it is usually necessary to choose one or other model as “ideal”, the same mixture of simulation, choice experiment and real-world application could be applied quite easily to assess any differences in the conclusions.

University of Cape Town

## Appendix A

# Supplementary material for Chapter 3

### A.1 Equivalence of quantile and fuzzy decision models

Let  $U_i^{(q_k)} = \sum_{j=1}^J w_j u_j(z_{ij}^{(q_k)})$  denote the evaluation of the utility of alternative  $a_i$  using quantile  $q_k$ , so that the fuzzy global utility  $\tilde{U}_i$  can also be written as the triangular fuzzy number  $[U_i^{(q_1)}, U_i^{(q_2)}, U_i^{(q_3)}]$  using (3.4). In the ranking method employed in [56], a discrete number of  $\alpha$ -cuts of  $\tilde{U}_i$  are taken, and at each  $\alpha$ -cut the two values of  $\tilde{U}_i$  at which  $\tilde{U}_i = \alpha$  are evaluated and used in the final crisp evaluation of  $a_i$  given by (3.5). Figure A.1 gives a graphical illustration of this approach for  $\alpha \in \{0, 0.25, 0.5, 0.75, 1\}$ . The contribution made to the final evaluation  $U_i^{(\text{fuzzy})}$  by a particular  $\alpha$ -cut is given by  $U_i^\alpha / \mathcal{N}$ .

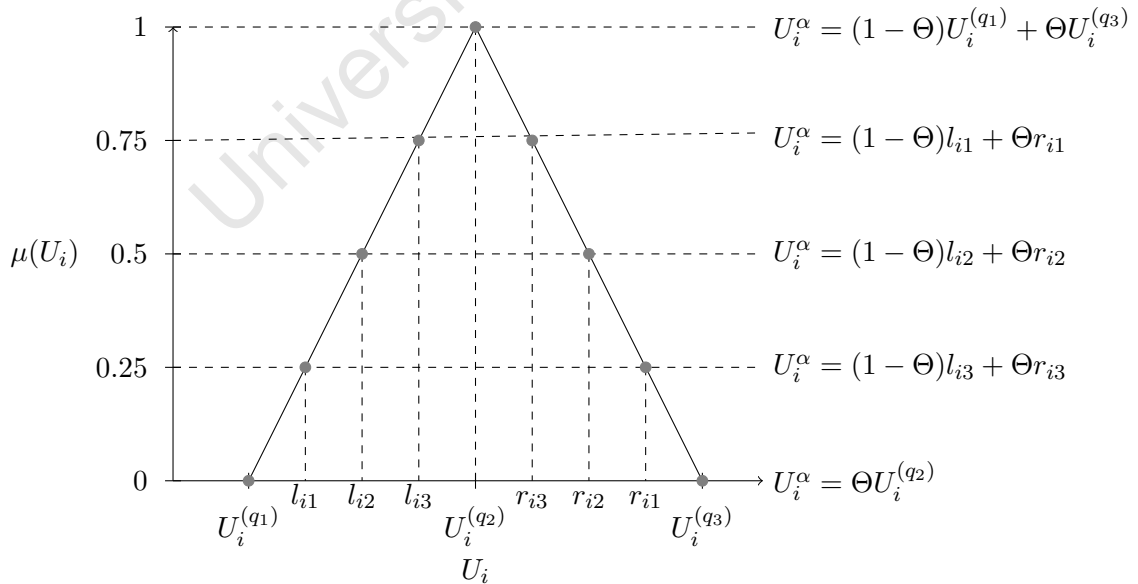


Figure A.1: Illustration of computations performed by fuzzy decision model

Figure A.1 shows how any value of  $U_i$  in the support of  $\tilde{U}_i$  (and thus any values that  $l_{in}$  and  $r_{in}$  might take on) can be represented by a weighted combination of the utility at the

three quantiles i.e.  $U_i^{(q_1)}$ ,  $U_i^{(q_2)}$ , and  $U_i^{(q_3)}$ . For example, for the case shown in Figure A.1,

$$\begin{aligned} l_{i1} &= 0.75U_i^{(q_1)} + 0.25U_i^{(q_2)} & r_{i1} &= 0.75U_i^{(q_3)} + 0.25U_i^{(q_2)} \\ l_{i2} &= 0.5U_i^{(q_1)} + 0.5U_i^{(q_2)} & r_{i2} &= 0.5U_i^{(q_3)} + 0.5U_i^{(q_2)} \\ l_{i3} &= 0.25U_i^{(q_1)} + 0.75U_i^{(q_2)} & r_{i3} &= 0.25U_i^{(q_3)} + 0.75U_i^{(q_2)} \end{aligned}$$

It follows that in general

$$U_i^{(\text{fuzzy})} = 0.5(1 - \Theta)U_i^{(q_1)} + 0.5U_i^{(q_2)} + 0.5\Theta U_i^{(q_3)}$$

Note that this equivalence only holds for the fuzzy decision model as it is defined here, and when triangular membership functions involving quantiles are used. Also note that in this case results are independent of the number of  $\alpha$ -cuts used,  $\mathcal{N}$ , and that the weight attached to the intermediate quantile is constrained to be equal to the sum of the other two weights (so that Keefer-Bodily or equal quantile weights, for example, are not possible). The formulation also suggests that the parameter  $\Theta$  should not be interpreted as an index of “optimism” (as it is in [56]) but as an indication of the relative importance of differences in performance at high and low quantiles.

## Appendix B

# Supplementary material for Chapter 4

### B.1 Documents given to study participants

Documents are shown on the following pages for one of the decision problems involving two objectives. The material comprises:

- An introduction to the problem, a summary of uncertainty formats, and survey questions to be answered following each task
- Attribute information for the decision problem
- A final survey presented at the end of the tasks

Thank you for participating in this study on decision making.

You will be required to make a number of investment decisions. The context of the decision, described below, will be the same for all the decisions. However, some aspects of the decision, like the number of investment objectives and the actual investment returns, will change from decision to decision.

Please carefully consider all information and make the best possible choice that you can, given the available information. You will have 4 minutes for each decision, and there are 12 decisions to be made.

### **The background**

A good friend of yours has received a bequest of £100 000 from a recently deceased relative. He wishes to invest the money, but knows nothing about the financial market. He has therefore asked you to invest the money for him. Depending on the problem, there may be one, two, or three investment objectives. Your friend has stated these to you as follows.

- OBJECTIVE 1: “After 1 year, I’m going to take out all the profit made in that time and take my wife on a holiday. The more money, the better the holiday we’ll take”.
- OBJECTIVE 2: “After 2 years, I’m going to take out all the profit again and buy a car. I think I’ll need at least £4000 to get any kind of decent car. Anything less than that and all I’ll be able to get is a piece of junk”.
- OBJECTIVE 3: “After 6 years, I want to take all the money and use it to buy a house”.

In addition, your friend has also told you:

- “Please put all my money into a single investment. I don’t want to have to keep track of more than this”.
- “Buying a house is more important to me than the holiday or the car, of course. The car and the holiday are about equally important”.

After some discussion with your friend you agree that the house is roughly twice as important as either the holiday or the car (i.e. in the ratio 2:1:1).



In this choice, there are TWO OBJECTIVES to consider and FIVE INVESTMENTS to choose from.

- OBJECTIVE 1: “After 1 year, I’m going to take out all the profit made in that time and take my wife on a holiday. The more money, the better the holiday we’ll take”.
- OBJECTIVE 2: “After 2 years, I’m going to take out all the profit again and buy a car. I think I’ll need at least £4000 to get any kind of decent car. Anything less than that and all I’ll be able to get is a piece of junk”.

Information in the form of SCENARIO ASSESSMENTS has been prepared by experts for you to consider in making your decision. Remember that each of these scenarios is considered equally likely to occur, but that it is also possible that none of the scenarios unfold *exactly* as predicted. See the attached note again if you are unfamiliar with the concept of scenarios.

Brief descriptions of each scenario are given below:

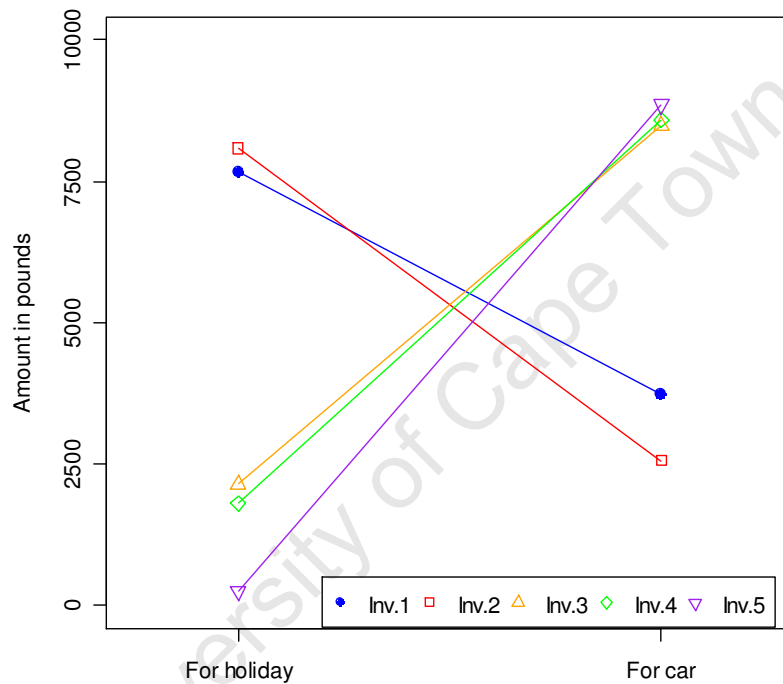
**The “Eastern Growth” Scenario:** In this scenario, countries in the Middle and Far East grow more rapidly in economic strength than the West (in particular, US and Europe). This growth, together with political stability, leads to good performance for investments with greater ties to the East.

**The “Nothing Changes” Scenario:** In this scenario, economic and political conditions remain relatively constant at their 2008 levels, for all major countries in which investments are held. The growth shown by China and India continues, but at a slower rate than previous experienced.

**The “Problematic World” Scenario:** In this scenario, some major economic shocks are experienced in several countries worldwide. Rising food prices leads to political instability in some of these countries. However, not all investments are affected and many investments, particularly in Western markets, still offer good performance.

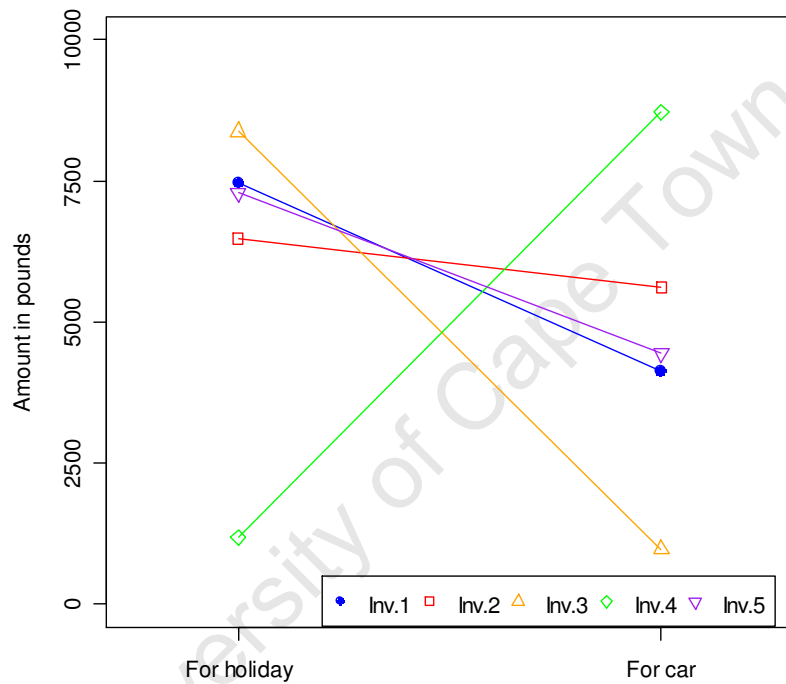
Investment performance in the “Eastern Growth” (EG) scenario

Investment	Amount available for holiday in EG scenario	Amount available for car in EG scenario
A	£ 7,667	£ 3,732
B	£ 8,093	£ 2,566
C	£ 2,151	£ 8,485
D	£ 1,806	£ 8,583
E	£ 247	£ 8,857



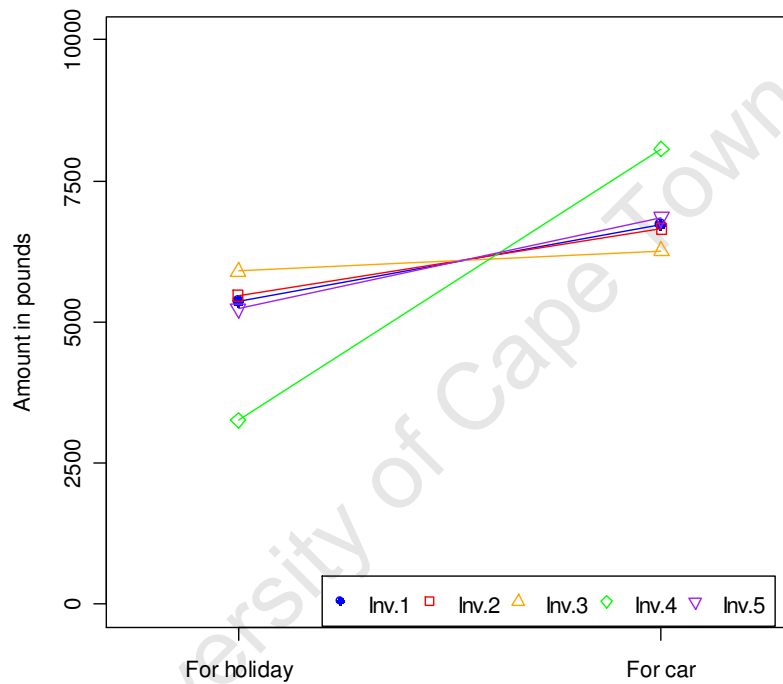
Investment performance in the “Nothing Changes” (NC) scenario

Investment	Amount available for holiday in NC scenario	Amount available for car in NC scenario
A	£ 7,469	£ 4,139
B	£ 6,484	£ 5,622
C	£ 8,376	£ 979
D	£ 1,193	£ 8,723
E	£ 7,296	£ 4,454



Investment performance in the “Problematic World” (PW) scenario

Investment	Amount available for holiday in PW scenario	Amount available for car in PW scenario
A	£ 5,375	£ 6,740
B	£ 5,469	£ 6,659
C	£ 5,902	£ 6,258
D	£ 3,265	£ 8,066
E	£ 5,235	£ 6,855



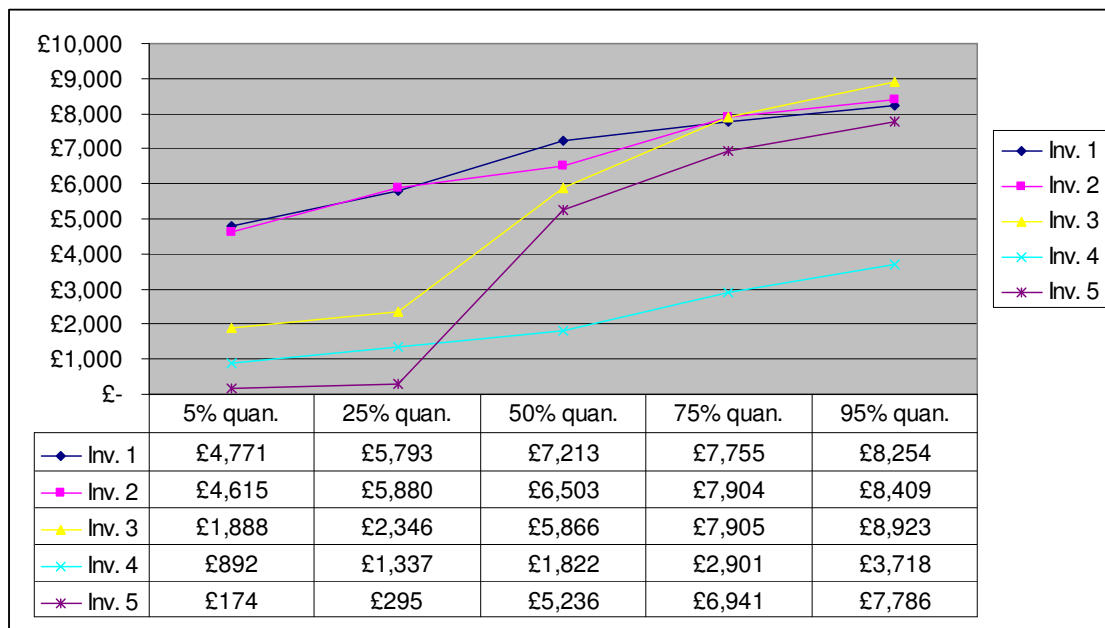
In this choice, there are TWO OBJECTIVES to consider and FIVE INVESTMENTS to choose from.

- OBJECTIVE 1: “After 1 year, I’m going to take out all the profit made in that time and take my wife on a holiday. The more money, the better the holiday we’ll take”.
- OBJECTIVE 2: “After 2 years, I’m going to take out all the profit again and buy a car. I think I’ll need at least £4000 to get any kind of decent car. Anything less than that and all I’ll be able to get is a piece of junk”.

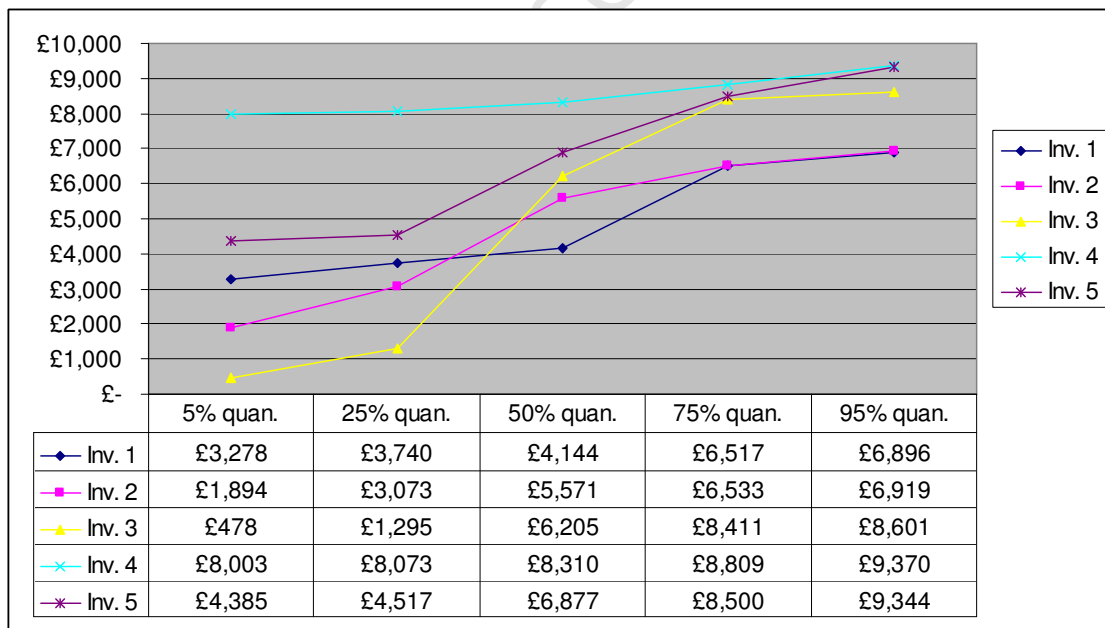
Information in the form of QUANTILES has been prepared by experts for you to consider in making your decision. See the attached note if you are unfamiliar with the concept of quantiles.

On the next page, the information is presented for you in both numerical and graphical formats. In the graph, each line represents one of the investments, with the monetary amounts plotted on the vertical ( $Y$ ) axis. Each objective has its own graph.

### Quantiles of amount available for holiday



### Quantiles of amount available for car



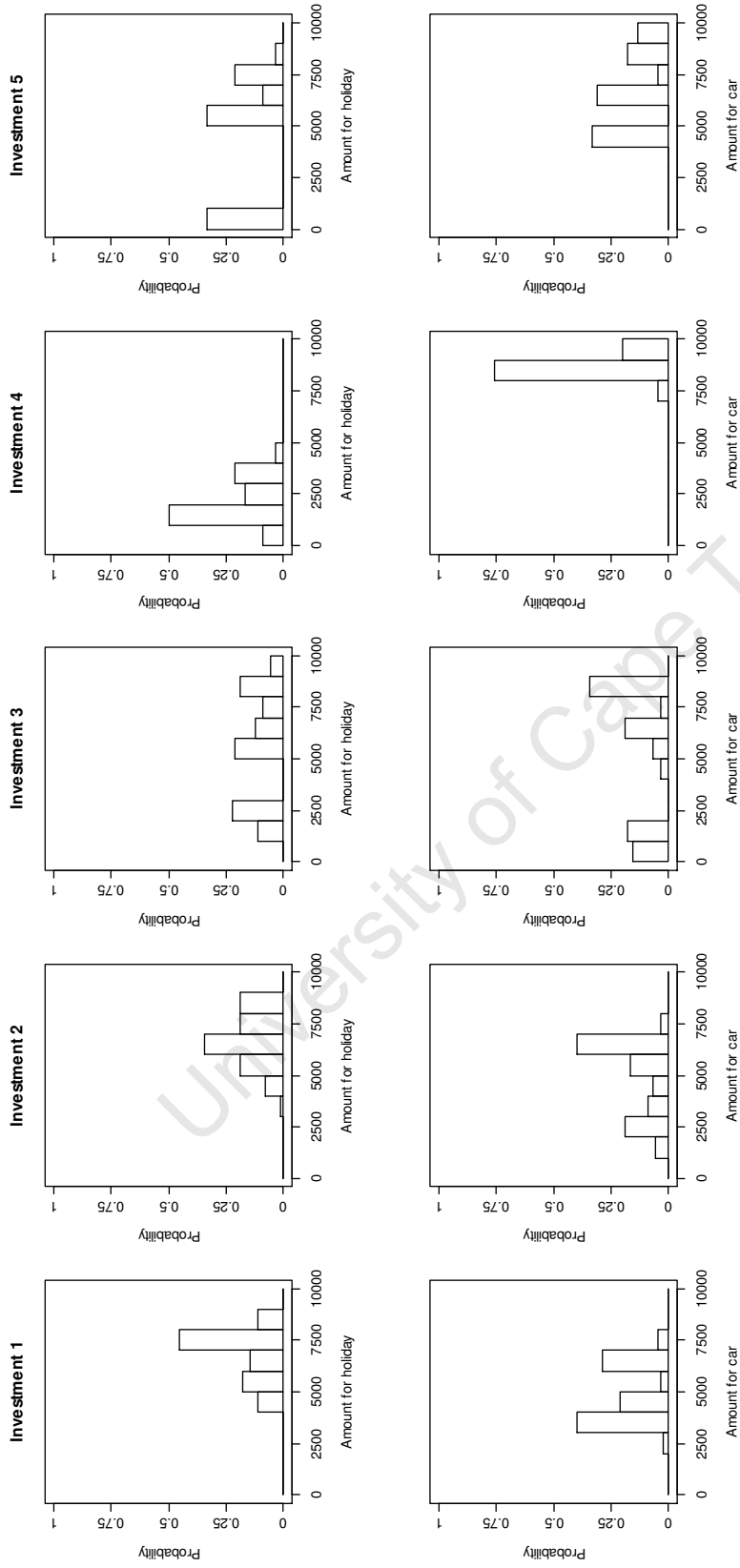
In this choice, there are TWO OBJECTIVES to consider and FIVE INVESTMENTS to choose from.

- OBJECTIVE 1: “After 1 year, I’m going to take out all the profit made in that time and take my wife on a holiday. The more money, the better the holiday we’ll take”.
- OBJECTIVE 2: “After 2 years, I’m going to take out all the profit again and buy a car. I think I’ll need at least £4000 to get any kind of decent car. Anything less than that and all I’ll be able to get is a piece of junk”.

Information in the form of PROBABILITY DISTRIBUTIONS has been prepared by experts for you to consider in making your decision. Read the attached note again if you are unfamiliar with the concept of probabilities.

On the next page, the information is presented for you in graphical format. The graphs are histograms, showing the probability that an investment returns amounts lying between certain intervals. On the X-axis, the investment returns have been divided into 10 intervals of £1000 each (£0 - £1000, £1000 - £2000, etc). The Y-axis specifies the probability that the investment return will lie within each of the given intervals.

**Important note:** graphs in the same column all refer to the same investment. Graphs in the same row refer to the same attribute.





In this choice, there are TWO OBJECTIVES to consider and FIVE INVESTMENTS to choose from.

- OBJECTIVE 1: “After 1 year, I’m going to take out all the profit made in that time and take my wife on a holiday. The more money, the better the holiday we’ll take”.
- OBJECTIVE 2: “After 2 years, I’m going to take out all the profit again and buy a car. I think I’ll need at least £4000 to get any kind of decent car. Anything less than that and all I’ll be able to get is a piece of junk”.

Information in the form of FUZZY NUMBERS has been prepared by experts for you to consider in making your decision. Read the attached note again if you are unfamiliar with the concept of fuzzy numbers.

On the next 2 pages, the information is presented for you in both numerical and graphical formats. The tables show three values, for each investment:

1. The minimum investment return that is considered possible.
2. The median investment return, which is considered to be “most representative” of the set of possible investment returns.
3. The maximum investment return that is considered possible

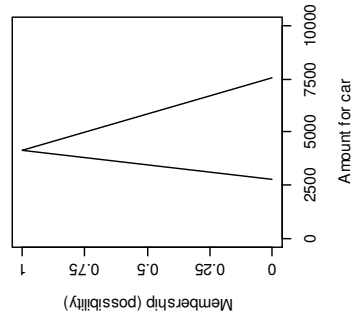
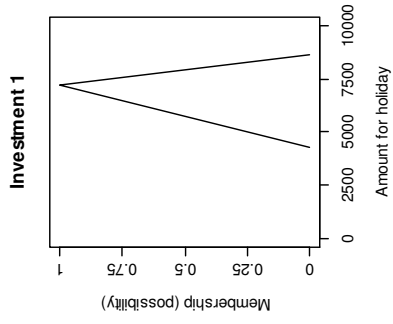
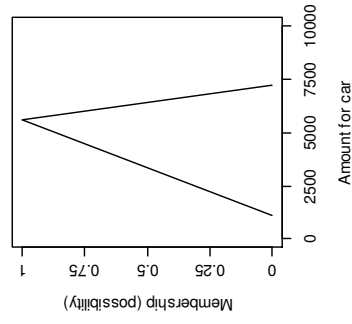
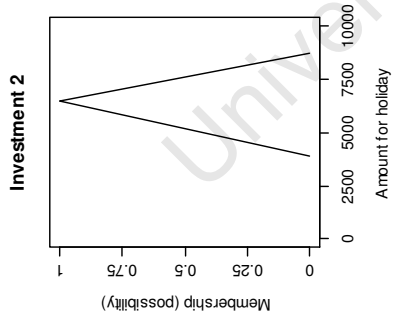
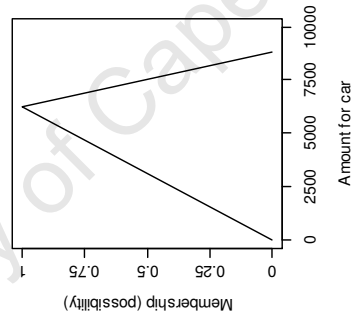
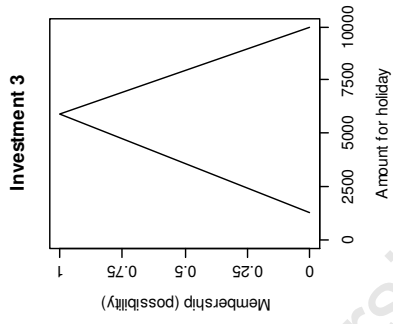
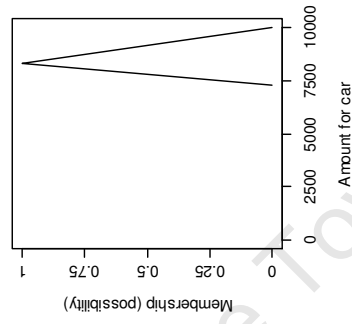
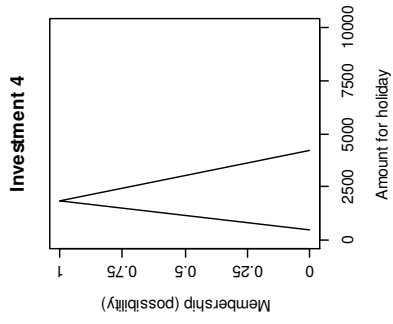
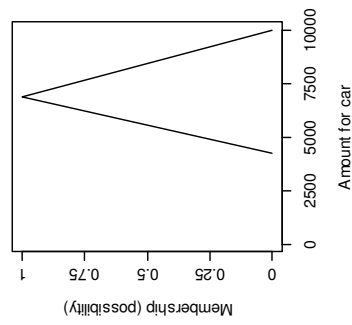
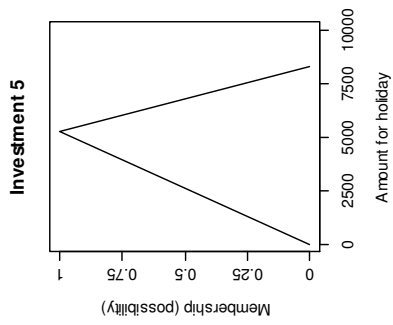
Note that there is one table for each attribute. The graphs simply plot the three values in the table for each investment, with linear interpolation being applied between the three values. All amounts are shown in pounds (£). **Important note:** graphs in the same column all refer to the same investment. Graphs in the same row refer to the same attribute.

Fuzzy investment returns for “Amount available for holiday”

<b>Investment</b>	<b>Minimum amount possible</b>	<b>Amount with largest membership</b>	<b>Maximum amount possible</b>
1	\$ 4,266	\$ 7,213	\$ 8,596
2	\$ 3,918	\$ 6,503	\$ 8,735
3	\$ 1,292	\$ 5,866	\$ 10,000
4	\$ 501	\$ 1,822	\$ 4,206
5	\$ 0	\$ 5,236	\$ 8,293

Fuzzy investment returns for “Amount available for car”

<b>Investment</b>	<b>Minimum amount possible</b>	<b>Amount with largest membership</b>	<b>Maximum amount possible</b>
1	\$ 2,802	\$ 4,144	\$ 7,555
2	\$ 1,088	\$ 5,571	\$ 7,209
3	\$ 0	\$ 6,205	\$ 8,847
4	\$ 7,300	\$ 8,310	\$ 9,978
5	\$ 4,278	\$ 6,877	\$ 10,000



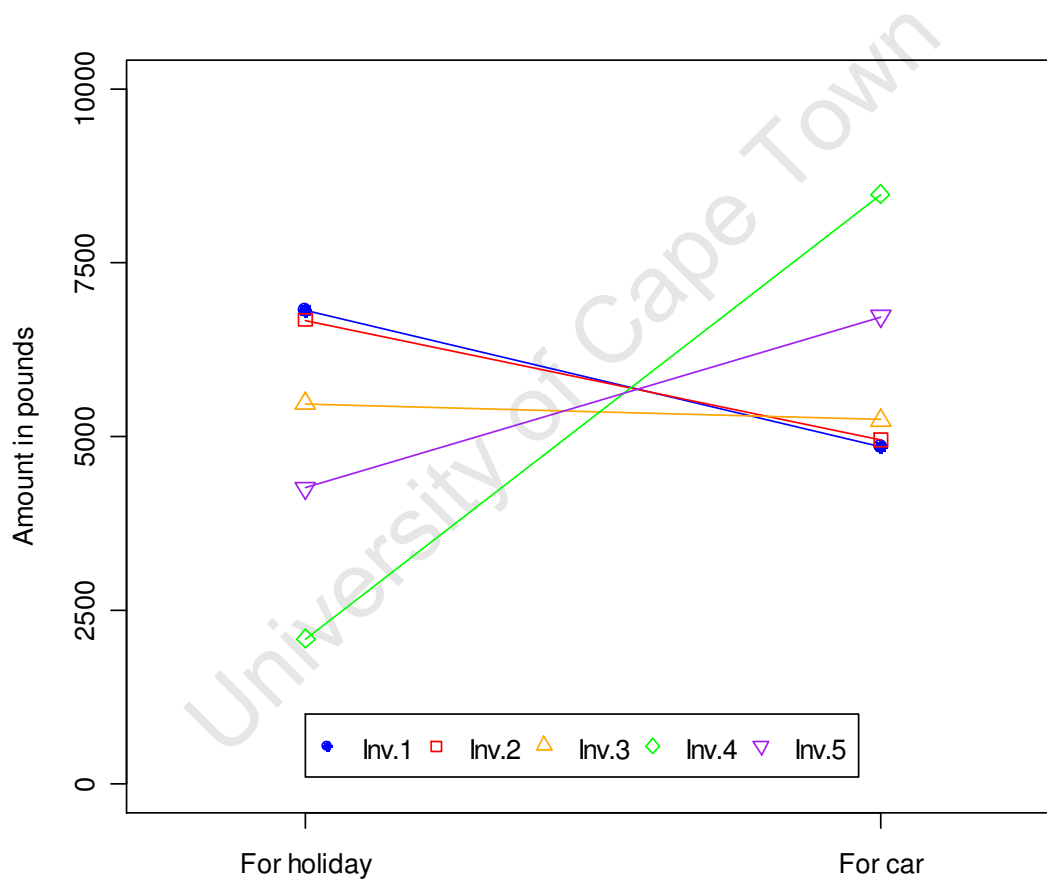
In this choice, there are TWO OBJECTIVES to consider and FIVE INVESTMENTS to choose from.

- OBJECTIVE 1: “After 1 year, I’m going to take out all the profit made in that time and take my wife on a holiday. The more money, the better the holiday we’ll take”.
- OBJECTIVE 2: “After 2 years, I’m going to take out all the profit again and buy a car. I think I’ll need at least £4000 to get any kind of decent car. Anything less than that and all I’ll be able to get is a piece of junk”.

Information in the form of EXPECTED VALUES has been prepared by experts for you to consider in making your decision. Remember that these are only *expected* values and not the amount your friend will receive *for sure* (see the attached note if you are unfamiliar with the concept of expected value).

On the next page, the information is presented for you in both numerical and graphical formats. In the graph, each line represents one of the investments, with the expected values plotted on the vertical (*Y*) axis.

Investment	Expected amount available for holiday	Expected amount available for car
1	£6,818	£4,855
2	£6,675	£4,949
3	£5,467	£5,233
4	£2,094	£8,479
5	£4,258	£6,724



In this choice, there are TWO OBJECTIVES to consider and FIVE INVESTMENTS to choose from.

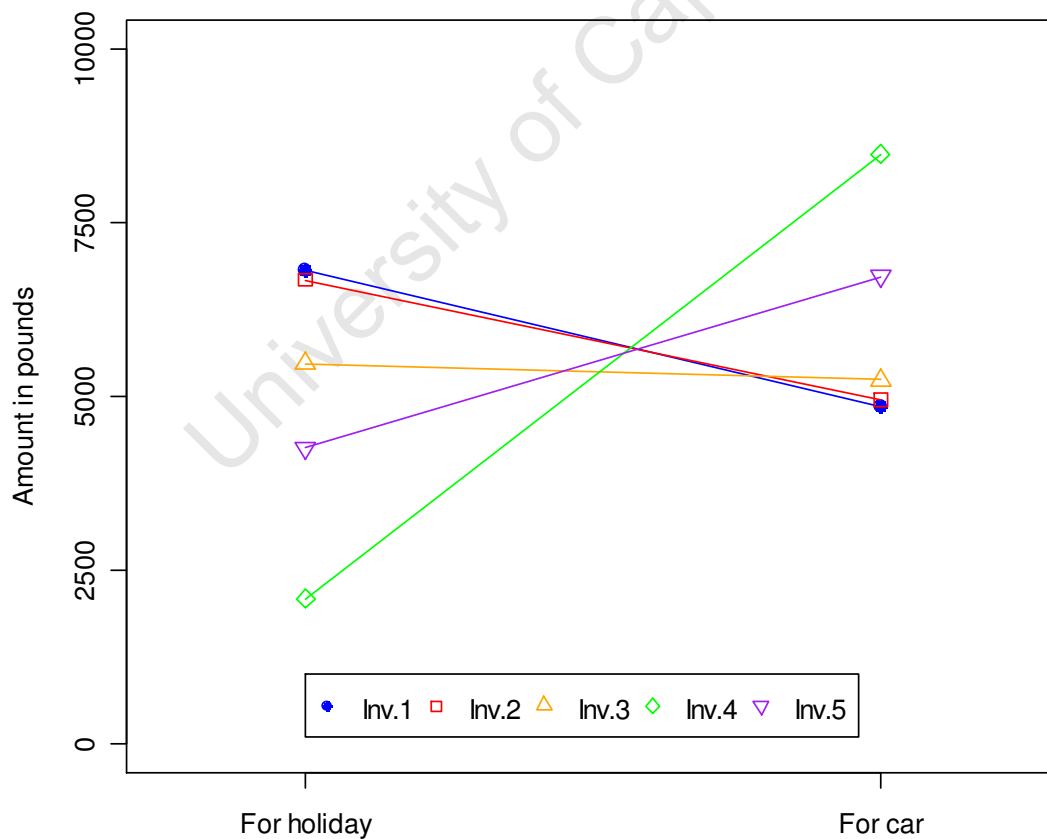
- OBJECTIVE 1: “After 1 year, I’m going to take out all the profit made in that time and take my wife on a holiday. The more money, the better the holiday we’ll take”.
- OBJECTIVE 2: “After 2 years, I’m going to take out all the profit again and buy a car. I think I’ll need at least £4000 to get any kind of decent car. Anything less than that and all I’ll be able to get is a piece of junk”.

Information in the form of EXPECTED VALUES and STANDARD DEVIATIONS has been prepared by experts for you to consider in making your decision. Remember that the expected values are only *expected* and not the amount your friend will receive *for sure* (see the attached note if you are unfamiliar with the concept of expected value).

On the next page, the expected values and standard deviations are presented for you in the form of a table. In addition, the expected values are shown in a graph below the tables. In the graph, each line represents one of the investments, with the expected values plotted on the vertical (*Y*) axis. **Important note:** the graph only shows the expected values i.e. does not include the standard deviations.

Investment	Expected amount available for holiday	Risk (standard deviation)
1	£6,818	£1,165
2	£6,675	£1,212
3	£5,467	£2,607
4	£2,094	£951
5	£4,258	£2,992

Investment	Expected amount available for car	Risk (standard deviation)
1	£4,855	£1,426
2	£4,949	£1,839
3	£5,233	£3,214
4	£8,479	£518
5	£6,724	£1,835



**TO BE ANSWERED AT THE END OF THE TASK**

Please answer the following four questions.

- When there was **ONE** objective, which way of presenting uncertainty made it **easier** for you **to make a decision**?

Probability distributions	Expected values	There was no difference
---------------------------	-----------------	-------------------------

- When there was **ONE** objective, which way of presenting uncertainty do you think helped you to make **better decisions**?

Probability distributions	Expected values	There was no difference
---------------------------	-----------------	-------------------------

- When there were **TWO** objectives, which way of presenting uncertainty made it **easier** for you **to make a decision**?

Probability distributions	Expected values	There was no difference
---------------------------	-----------------	-------------------------

- When there were **TWO** objectives, which way of presenting uncertainty do you think helped you to make **better decisions**?

Probability distributions	Expected values	There was no difference
---------------------------	-----------------	-------------------------



## B.2 Decision models used

In the following, alternatives are denoted by  $a_i$ ,  $i \in \{1, \dots, I\}$ , and attributes by  $c_j$ ,  $j \in \{1, \dots, J\}$ . The (stochastic) attribute evaluation of  $a_i$  on  $c_j$  is denoted by  $Z_{ij}$ , and  $u_j(Z_{ij})$  are marginal utility functions.

### B.2.1 Additive MAUT model

The evaluation of  $a_i$  is given by  $U_i = \sum_{j=1}^J w_j E[u_j(Z_{ij})]$  where  $U_i$  is the expected utility of alternative  $a_i$  and  $w_j$  is an attribute importance weight indicating the relative importance of a one-unit change in attribute  $c_j$  [e.g. 31].

### B.2.2 Expected values

The evaluation of  $a_i$  is given by  $U_i^{(\text{ev})} = \sum_{j=1}^J w_j u_j(E[Z_{ij}])$ .

### B.2.3 Expected values and standard deviations

The evaluation of  $a_i$  is given by  $U_i^{(\text{risk})} = \sum_{j=1}^J w_j u_j(E[Z_{ij}]) - w_{ij}^R \sigma_{ij}^2$ , where  $\sigma_{ij}^2$  is the variance of  $Z_{ij}$  and  $w_{ij}^R$  is a ‘risk’ weight for  $\sigma_{ij}^2$ . Risk weights are not specified by the problem description, so it is not possible to identify idealised weights. In calculating utility loss relative to an idealised application of standard deviations, I use eight different risk weights,  $w_{ij}^R \in \{0.25, 0.5, 0.75, 1, 2, 3, 4\}$ , and select the minimum utility loss obtained i.e. the utility loss showing the decision maker in the most favourable light.

### B.2.4 Quantiles

The evaluation of  $a_i$  is given by  $U_i^{(\text{quan})} = \sum_{k=1}^{N_q} [w_{q_k} \sum_{j=1}^J w_j u_j(z_{ij}^{(q_k)})]$ , where  $q_k$  refers to a specific quantile,  $z_{ij}^{(q_k)}$  is the  $q_k$ -th quantile of  $Z_{ij}$ ,  $w_{q_k}$  denotes the weight associated with quantile  $q_k$ , and  $N_q$  is the number of quantiles used. I use five quantiles i.e.  $N_q = 5$ ,  $\mathbf{q} = (0.05, 0.25, 0.5, 0.75, 0.95)$ ; and equal weights for each quantile i.e.  $w_q = 1/N_q$ .

### B.2.5 Three-point approximations/fuzzy numbers

The fuzzy global evaluations are computed using standard fuzzy operations [e.g 85] as  $\tilde{U}_i = [\sum_{j=1}^J w_j u_j(z_{ij}^{(0)}), \sum_{j=1}^J w_j u_j(z_{ij}^{(0.5)}), \sum_{j=1}^J w_j u_j(z_{ij}^{(1)})]$ , where  $z_{ij}^{(0)}$ ,  $z_{ij}^{(0.5)}$  and  $z_{ij}^{(1)}$  denote the minimum, median, and maximum of  $Z_{ij}$  respectively. These evaluations can be ranked by any number of fuzzy ranking procedures – this chapter employs an approach based on left and right dominance [56], which is itself a generalization of the area compensation approach in [99]. A final crisp score on which the alternatives can be unambiguously ranked is given by  $U_i^{(\text{fuzzy})} = \frac{1}{N+1} [\Theta \sum_{n=0}^N r_{in} + (1 - \Theta) \sum_{n=0}^N l_{in}]$  where  $r_{in}$  and  $l_{in}$  are the upper and lower limits of the  $n^{\text{th}}$   $\alpha$ -cut of the fuzzy number  $\tilde{U}_i$  (so-called right- and left-dominance) respectively, and  $\Theta \in [0, 1]$  is the weight assigned to right-dominance. The

$\alpha$ -cuts used are given by  $\alpha_{\tilde{c}} = n/\mathcal{N}$  for  $n = \{0, 1, \dots, \mathcal{N}\}$ . I use  $\mathcal{N} = 6$  on the basis of results in [56] indicating little change in results beyond this value, and an equal weighting for right- and left-dominance i.e.  $\Theta = 0.5$ , although conclusions are not sensitive to choices of either parameter.

### B.2.6 Scenarios

The evaluation of  $a_i$  is given by  $U_i^{(\text{scen})} = \sum_{k=1}^{N_s} [w_{s_k} \sum_{j=1}^J w_j u_j(\mu_{ijk})]$  where  $s_k$  refers to a specific scenario,  $\mu_{ijk}$  is the mean evaluation of alternative  $a_i$  on attribute  $c_j$  in broad future  $k$ ,  $w_{s_k}$  is the weight associated with scenario  $s_k$ , and  $N_s$  is the number of scenarios used. I use three scenarios i.e.  $N_s = 3$ , and equal scenario weights i.e.  $w_{s_k} = 1/N_s$ .

## B.3 Utility function parameters used

The standardised exponential form is again used to represent marginal utility functions:

$$u_j(x) = \begin{cases} \frac{\lambda_j(e^{\alpha_j x} - 1)}{e^{\alpha_j \tau_j} - 1} & \text{for } 0 \leq x \leq \tau_j \\ \lambda_j + \frac{(1 - \lambda_j)(1 - e^{-\beta_j(x - \tau_j)})}{1 - e^{-\beta_j(1 - \tau_j)}} & \text{for } \tau_j < x \leq 1 \end{cases}$$

with parameters given in Table B.1.

Set	Attribute 1				Attribute 2				Attribute 3			
	$\tau_1$	$\lambda_1$	$\alpha_1$	$\beta_1$	$\tau_2$	$\lambda_2$	$\alpha_2$	$\beta_2$	$\tau_3$	$\lambda_3$	$\alpha_3$	$\beta_3$
1	0	0	0	0	0	0	0.002	0	90	0	0.002	0
2	0	0	4	2	0	0	4	2	90	0	4	2
3	0	0	8	4	0	0	8	4	90	0	8	4
4	0	0	12	6	0	0	32	16	90	0	12	6
5	0	0	0	0	4	0.9	0.002	0	90	0	0.002	0
6	0	0	4	2	4	0.9	4	2	90	0	4	2
7	0	0	8	4	4	0.9	8	4	90	0	8	4
8	0	0	12	6	4	0.9	32	16	90	0	12	6
9	0	0	0	0	4	0.9	0.002	0	100	0.6	0.002	0
10	0	0	4	2	4	0.9	4	2	100	0.6	4	2
11	0	0	8	4	4	0.9	8	4	100	0.6	8	4
12	0	0	12	6	4	0.9	32	16	100	0.6	12	6

Table B.1: Parameter estimates for utility functions used to compute utility loss. Reference point values ( $\tau_j$ ) are given in thousands of pounds.

## B.4 Accuracy results for all utility functions

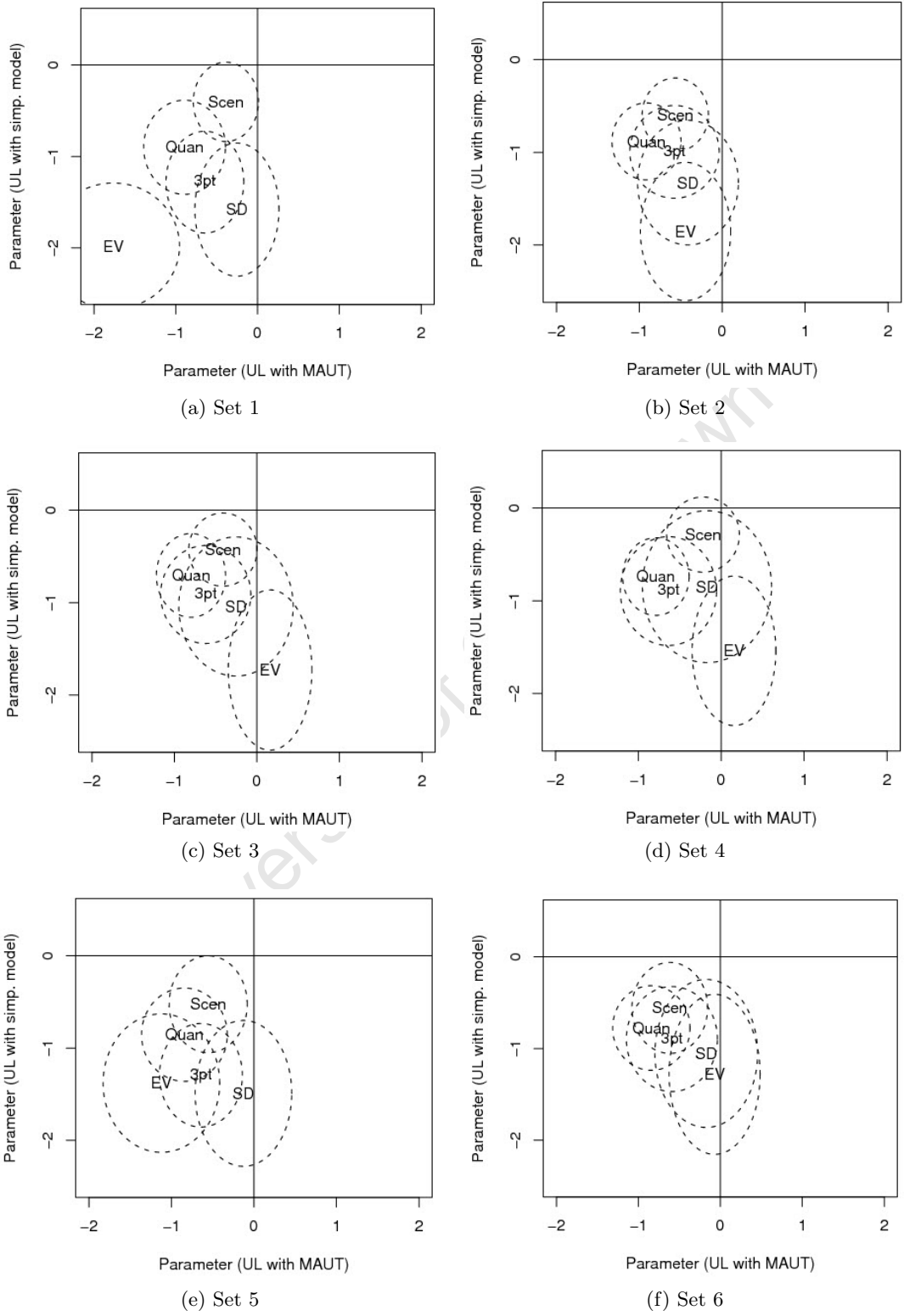


Figure B.1: Estimates of the effects of uncertainty format on accuracy using different utility functions (Sets 1 to 6).

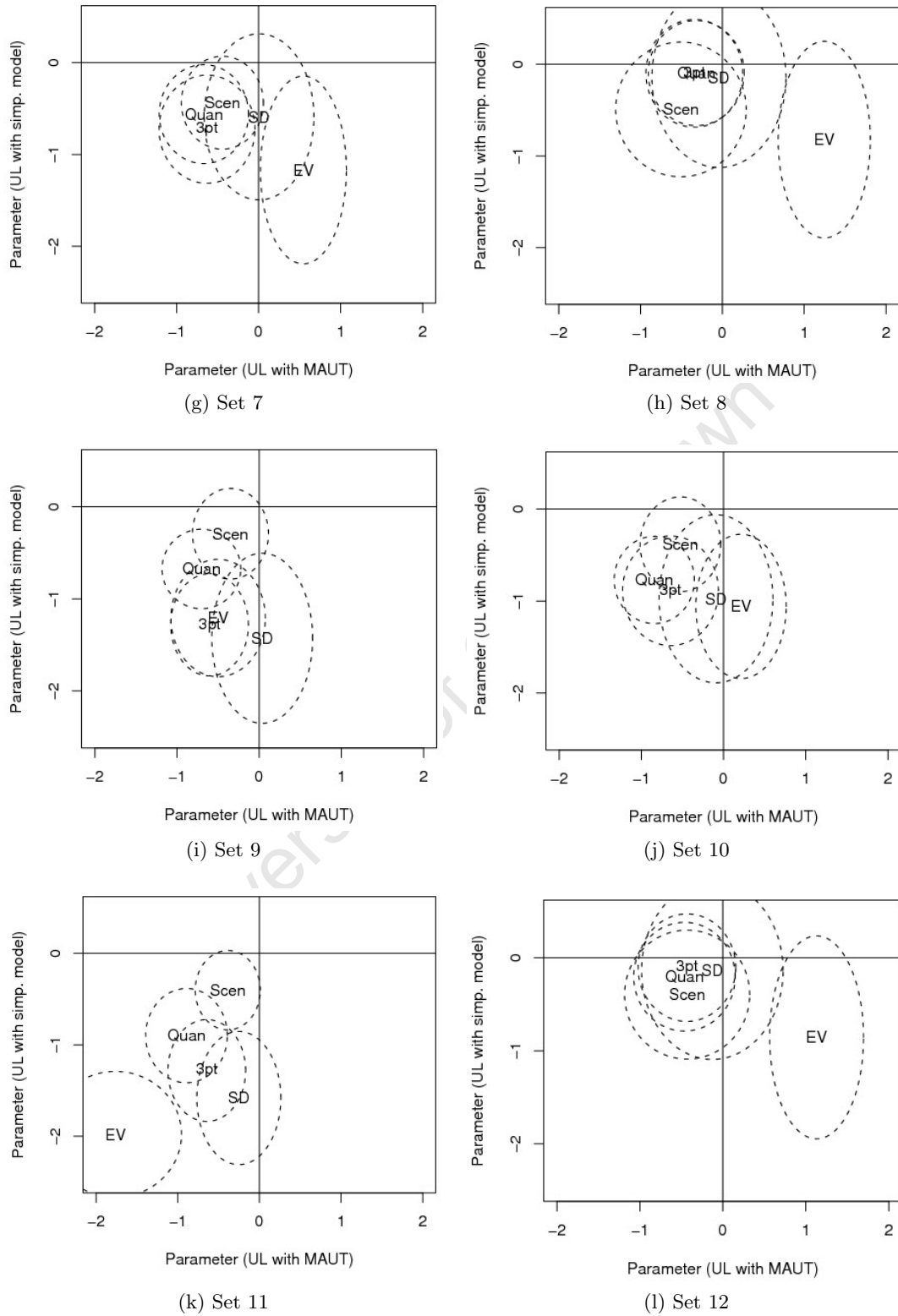


Figure B.1: Estimates of the effects of uncertainty format on accuracy using different utility functions (Sets 7 to 12).

# Bibliography

- [1] S. Abbasbandy and B. Asady. Ranking of fuzzy numbers by sign distance. *Information Sciences*, 176(16):2405–2416, 2006.
- [2] S. Abbasbandy and T. Hajjari. A new approach for ranking of trapezoidal fuzzy numbers. *Computers and Mathematics with Applications*, 57:413–419, 2009.
- [3] M. Abdellaoui. Parameter-free elicitation of utility and probability weighting functions. *Management Science*, pages 1497–1512, 2000.
- [4] M. Abdellaoui, C. Barrios, and P. Wakker. Reconciling introspective utility with revealed preference: Experimental arguments based on prospect theory. *Journal of Econometrics*, 138(1):356–378, 2007.
- [5] M. Abdellaoui, H. Bleichrodt, and C. Paraschiv. Loss aversion under prospect theory: A parameter-free measurement. *Management Science*, 53(10):1659–1674, 2007.
- [6] W. Abrahamse, L. Steg, C. Vlek, and T. Rothengatter. A review of intervention studies aimed at household energy conservation. *Journal of Environmental Psychology*, 25(3):273–291, 2005.
- [7] 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.
- [8] M. Alpert and H. Raiffa. A progress report on the training of probability assessors. In *Judgment under uncertainty: Heuristics and biases* [138], pages 306–334.
- [9] P. Anand. *Foundations of rational choice under risk*. Clarendon Press, 1993.
- [10] 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.
- [11] B. Asady and A. Zendehnam. Ranking fuzzy numbers by distance minimization. *Applied Mathematical Modelling*, 31(11):2589–2598, 2007.

- [12] B. Ashtiani, F. Haghghirad, A. Makui, and G. ali Montazer. Extension of fuzzy TOPSIS method based on interval-valued fuzzy sets. *Applied Soft Computing*, 9:457–461, 2009.
- [13] R. Ashton. Effects of justification and a mechanical aid on judgment performance. *Organizational Behavior and Human Decision Processes*, 52(2):292–306, 1992.
- [14] S. Azondékon and J. Martel. Value of additional information in multicriterion analysis under uncertainty. *European Journal of Operational Research*, 117:45–62, 1999.
- [15] E. Balletero. A compromise approach to specification and optimization. *Journal of Multi-criteria Decision Analysis*, 6:11–16, 1997.
- [16] E. Balletero. Stochastic goal programming: A mean-variance approach. *European Journal of Operational Research*, 131:476–481, 2001.
- [17] C. Bana e Costa. A multicriteria decision aid methodology to deal with conflicting situations on the weights. *European Journal of Operational Research*, 26(1):22–34, 1986.
- [18] R. Banuelas and J. Antony. Application of stochastic analytic hierarchy process within a domestic appliance manufacturer. *Journal of the Operational Research Society*, 58(1):29, 2007.
- [19] F. Barron. Influence of missing attributes on selecting a best multiattributed alternative. *Decision Sciences*, 18:194–205, 1987.
- [20] F. Barron and B. Barrett. Decision quality using ranked attribute weights. *Management Science*, 42(11):1515–1523, 1996.
- [21] J. Barthélemy and E. Mullet. Choice basis: A model for multi-attribute preference. *British journal of mathematical & statistical psychology*, 39(1):106–124, 1986.
- [22] I. Basak. Estimation of the multi-criteria worths of the alternatives in a hierarchical structure of comparisons. *Communications in Statistics-Theory and Methods*, 18(10):3719–3738, 1989.
- [23] I. Basak. Inference in pairwise comparison experiments based on ratio scales. *Journal of Mathematical Psychology*, 35(1):80–91, 1991.
- [24] I. Basak. Probabilistic judgments specified partially in the analytic hierarchy process. *European Journal of Operational Research*, 108(1):153–164, 1998.
- [25] V. Bawa. Optimal rules for ordering uncertain prospects. *Journal of Financial Economics*, 2(1):95–121, 1975.

- [26] L. Beach, J. Christensen-Szalanski, and V. Barnes. Assessing human judgment: Has it been done, can it be done, should it be done. In Wright and Ayton [287], pages 49–62.
- [27] L. Beach and T. Scopp. Intuitive statistical inferences about variances. *Organizational Behavior and Human Performance*, 3(2):109–123, 1968.
- [28] L. Beach and R. Swenson. Intuitive estimation of means. *Psychonomic Science*, 5(4):161–162, 1966.
- [29] J. Becker and R. Sarin. Gamble Dependent Utility. *Management Science*, 33:1367–1382, 1987.
- [30] D. Bell. Disappointment in Decision Making under Uncertainty. *Operations Research*, 33(1):1–27, 1985.
- [31] V. Belton and T. Stewart. *Multiple Criteria Decision Analysis: An Integrated Approach*. Kluwer Academic Publishers, Boston, 2002.
- [32] H. Bleichrodt and J. Miyamoto. A characterization of quality-adjusted life-years under cumulative prospect theory. *Mathematics of Operations Research*, pages 181–193, 2003.
- [33] H. Bleichrodt and J. Pinto. A parameter-free elicitation of the probability weighting function in medical decision analysis. *Management Science*, pages 1485–1496, 2000.
- [34] H. Bleichrodt, J. Pinto, and P. Wakker. Making descriptive use of prospect theory to improve the prescriptive use of expected utility. *Management Science*, pages 1498–1514, 2001.
- [35] C. Boender, J. de Graan, and F. Lootsma. Multicriteria decision analysis with fuzzy pairwise comparisons. *Fuzzy Sets and Systems*, 29(2):133–143, 1989.
- [36] F. Bolger and G. Wright. Assessing the quality of expert judgment: Issues and analysis. *Decision support systems*, 11(1):1–24, 1994.
- [37] R. Bolton and J. Drew. A longitudinal analysis of the impact of service changes on customer attitudes. *Journal of Marketing*, 55(1):1–9, 1991.
- [38] R. Bordley and C. Kirkwood. Multiattribute Preference Analysis with Performance Targets. *Operations Research*, 52(6):823, 2004.
- [39] M. Bose Anne et al. Multi-attribute utility methods in group decision making: past applications and potential for inclusion in GDSS. *Omega*, 25(6):691–706, 1997.

- [40] A. Bryman and E. Bell. *Business research methods*. Oxford University Press, USA, 2007.
- [41] J. Buckley. Fuzzy hierarchical analysis. *Fuzzy Sets and Systems*, 17(3):233–247, 1985.
- [42] D. Bunn and A. Salo. Forecasting with scenarios. *European Journal of Operational Research*, 68:291–303, 1993.
- [43] J. Butler, J. Jia, and J. Dyer. Simulation techniques for the sensitivity of multi-criteria decision models. *European Journal of Operational Research*, 103:531–546, 1997.
- [44] G. Cairns, G. Wright, R. Bradfield, K. van der Heijden, and G. Burt. Exploring e-government futures through the application of scenario planning. *Technological Forecasting & Social Change*, 71(3):217–238, 2004.
- [45] C. Camerer. An experimental test of several generalized utility theories. *Journal of Risk and Uncertainty*, 2(1):61–104, 1989.
- [46] C. Camerer. Recent tests of generalizations of expected utility theory. In W. Edwards, editor, *Utility: Theories, Measurement and Applications*, pages 207–251. Kluwer, Boston, 1992.
- [47] C. Camerer and T. Ho. Violations of the betweenness axiom and nonlinearity in probability. *Journal of Risk and Uncertainty*, 8(2):167–196, 1994.
- [48] E. Carbone and J. Hey. Estimation of expected utility and non-expected utility preference functionals using complete ranking data. In Munier and Machina [200], pages 119–140.
- [49] E. Castagnoli and M. Calzi. Expected utility without utility. *Theory and Decision*, 41(3):281–301, 1996.
- [50] D. Chang. Applications of the extent analysis method on fuzzy AHP. *European Journal of Operational Research*, 95(3):649–655, 1996.
- [51] T.-H. Chang and T.-C. Wang. Using the fuzzy multi-criteria decision making approach for measuring the possibility of successful knowledge management. *Information Sciences*, 179:355–370, 2009.
- [52] Y. Chantal, R. Vallerand, and E. Vallieres. Motivation and gambling involvement. *The Journal of social psychology*, 135(6):755–763, 1995.



- [53] C. Chen. Extensions of the TOPSIS for group decision-making under fuzzy environment. *Fuzzy Sets and Systems*, 114(1):1–9, 2000.
- [54] C. Chen and C. Klein. An efficient approach to solving fuzzy MADM problems. *Fuzzy Sets and Systems*, 88(1):51–67, 1997.
- [55] C. Chen and H. Tang. Ranking nonnormal  $p$ -norm trapezoidal fuzzy numbers with integral value. *Computers and Mathematics with Applications*, 56:2340–2346, 2008.
- [56] L. Chen and H. Lu. An approximate approach for ranking fuzzy numbers based on left and right dominance. *Computers and Mathematics with Applications*, 41(12):1589–1602, 2001.
- [57] S. Chen and S. Chen. Fuzzy risk analysis based on similarity measures of generalized fuzzy numbers. *IEEE Transactions on Fuzzy Systems*, 11(1):45–56, 2003.
- [58] S. Chen and S. Chen. Fuzzy risk analysis based on the ranking of generalized trapezoidal fuzzy numbers. *Applied Intelligence*, 26(1):1–11, 2007.
- [59] S.-J. Chen and C.-L. Hwang. *Fuzzy Multiple Attribute Decision Making: Methods and Applications*. Springer-Verlag, Berlin, 1992.
- [60] S.-M. Chen and J.-H. Chen. Fuzzy risk analysis based on ranking generalized fuzzy numbers with different heights and different spreads. *Expert Systems With Applications*, 36:6833–6842, 2009.
- [61] S.-M. Chen and C.-H. Wang. Fuzzy risk analysis based on ranking fuzzy numbers using  $\alpha$ -cuts, belief features and signal/noise ratios. *Expert Systems With Applications*, 36:5576–5581, 2009.
- [62] C. Cheng. A new approach for ranking fuzzy numbers by distance method. *Fuzzy Sets and Systems*, 95(3):307–317, 1998.
- [63] T. Chermack. Improving decision-making with scenario planning. *Futures*, 36(3):295–309, 2004.
- [64] S. Chew. A Generalization of the Quasilinear Mean with Applications to the Measurement of Income Inequality and Decision Theory Resolving the Allais Paradox. *Econometrica*, 51(4):1065–1092, 1983.
- [65] S. Chew. Axiomatic utility theories with the betweenness property. *Annals of Operations Research*, 19(1):273–298, 1989.
- [66] S. Chew, L. Epstein, and U. Segal. Mixture Symmetry and Quadratic Utility. *Econometrica*, 59(1):139–163, 1991.

- [67] S. Chew and W. Waller. Empirical Tests of Weighted Utility Theory. *Journal of Mathematical Psychology*, 30(1):55–72, 1986.
- [68] T. Chu and Y. Lin. A Fuzzy TOPSIS Method for Robot Selection. *The International Journal of Advanced Manufacturing Technology*, 21(4):284–290, 2003.
- [69] T. Chu and C. Tsao. Ranking fuzzy numbers with an area between the centroid point and original point. *Computers and Mathematics with Applications*, 43(1-2):111–117, 2002.
- [70] T.-C. Chu and Y. Lin. An extension to fuzzy MCDM. *Computers and Mathematics with Applications*, 57:445–454, 2009.
- [71] J. Conlisk. Three Variants on the Allais Example. *American Economic Review*, 79(3):392–407, 1989.
- [72] R. Csutora and J. Buckley. Fuzzy hierarchical analysis: the Lambda-Max method. *Fuzzy Sets and Systems*, 120(2):181–195, 2001.
- [73] G. d’Avignon and P. Vincke. An outranking method under uncertainty. *European Journal of Operational Research*, 36:311–321, 1988.
- [74] R. Dawes. The robust beauty of improper linear models in decision making. *American Psychologist*, 34(7):571–582, 1979.
- [75] 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.
- [76] K. De Wulf, G. Odekerken-Schröder, and D. Iacobucci. Investments in consumer relationships: a cross-country and cross-industry exploration. *Journal of Marketing*, 65(4):33–50, 2001.
- [77] E. Dekel. An Axiomatic Characterization of Preferences under Uncertainty: Weakening the Independence Axiom. *Journal of Economic Theory*, 40(2):304–318, 1986.
- [78] B. Dendrou, S. Dendrou, and E. Houstis. Multiobjective decision analysis for engineering systems. *Computers and Operations Research*, 7:301–312, 1980.
- [79] H. Deng. Multicriteria analysis with fuzzy pairwise comparison. *International Journal of Approximate Reasoning*, 21:215–231, 1999.
- [80] Y. Deng and Q. Liu. A TOPSIS-based centroid-index ranking method of fuzzy numbers and its application in decision-making. *Cybernetics and Systems*, 36(6):581–595, 2005.

- [81] Y. Deng, Z. Zhenfu, and Q. Liu. Ranking fuzzy numbers with an area method using radius of gyration. *Computers and Mathematics with Applications*, 51(6-7):1127–1136, 2006.
- [82] D. Dey and J. Liu. A Quantitative Study of Quantile Based Direct Prior Elicitation from Expert Opinion. *Bayesian Analysis*, 2(1):137–166, 2007.
- [83] R. Donckels. Regional multiobjective planning under uncertainty: A stochastic goal programming formulation. *Journal of Regional Science*, 17(2):207–216, 1977.
- [84] N. Dowling, D. Smith, and T. Thomas. Electronic gaming machines: are they the ‘crack-cocaine’ of gambling? *Addiction*, 100:33–45, 2005.
- [85] D. Dubois and H. Prade. *Fuzzy Sets and Systems: Theory and Applications*. Academic Press New York, 1980.
- [86] M. Dumont and R. Ladouceur. Evaluation of motivation among video-poker players. *Psychological Reports*, 66:95–98, 1990.
- [87] I. Durbach. A simulation-based test of stochastic multicriteria acceptability analysis using achievement functions. *European Journal of Operational Research*, 170:923–934, 2006.
- [88] I. Durbach. On the estimation of a satisficing model of choice using stochastic multicriteria acceptability analysis. *Omega*, 37(3):497–509, 2009.
- [89] I. Durbach. The use of the SMAA acceptability index in descriptive decision analysis. *European Journal of Operational Research*, 37(196):1229–1237, 2009.
- [90] I. Durbach and T. Stewart. Integrating scenario planning and goal programming. *Journal of Multi-Criteria Decision Analysis*, 12(4-5):261–271, 2003.
- [91] I. Durbach and T. Stewart. Using expected values to simplify decision making under uncertainty. *Omega*, 37(2):312–330, 2009.
- [92] R. Dyckerhoff. Decomposition of multivariate utility functions in non-additive expected utility theory. *Journal of Multi-Criteria Decision Analysis*, 3(1), 1994.
- [93] A. Ehrenberg. *Repeat-buying: facts, theory and applications*. Charles Griffin & Company Limited, London, 1988.
- [94] D. Ellsberg. Risk, Ambiguity, and the Savage Axioms. *Quarterly Journal of Economics*, 75:643–669, 1961.
- [95] B. Enserink. Building scenarios for the University. *International Transactions in Operational Research*, 7(6):569–583, 2000.

- [96] N. Fenton and W. Wang. Risk and confidence analysis for fuzzy multicriteria decision making. *Knowledge-Based Systems*, 19(6):430–437, 2006.
- [97] B. Fischhoff, W. de Bruin, U. Güvenc, D. Caruso, and L. Brilliant. Analyzing disaster risks and plans: An avian flu example. *Journal of Risk and Uncertainty*, 33:131–149, 2006.
- [98] P. Fishburn. Foundations of Risk Measurement. I. Risk as probable loss. *Management Science*, 30(4):396–405, 1984.
- [99] P. Fortemps, M. Roubens, and I. Liege. Université de. Ranking and defuzzification methods based on area compensation. *Fuzzy Sets and Systems*, 82(3):319–330, 1996.
- [100] S. French. *Decision Theory: An Introduction to the Mathematics of Rationality*. Ellis Horwood, Chichester, 1988.
- [101] S. French. Uncertainty and imprecision: Modelling and analysis. *Journal of the Operational Research Society*, 46:70–79, 1995.
- [102] P. Garthwaite, J. Kadane, and A. O’Hagan. Statistical methods for eliciting probability distributions. *Journal of the American Statistical Association*, 100(470):680–701, 2005.
- [103] P. Garthwaite and A. O’Hagan. Quantifying expert opinion in the UK water industry: an experimental study. *The Statistician*, pages 455–477, 2000.
- [104] T. Gilovich. *How we know what isn’t so: The fallibility of human reason in everyday life*. Free Press, New York, 1991.
- [105] T. Gilovich, D. Griffin, and D. Kahneman. *Heuristics and biases: The psychology of intuitive judgment*. Cambridge University Press, 2002.
- [106] G. Goodhardt, A. Ehrenberg, and C. Chatfield. The Dirichlet: a comprehensive model of buying behaviour. *Journal of the Royal Statistical Society. Series A (General)*, 147(5):621–655, 1984.
- [107] J. Goodman and D. Ward. The importance of customer satisfaction. *Direct Marketing*, 56(8):23–6, 1993.
- [108] P. Goodwin and G. Wright. *Decision Analysis for Management Judgement*. John Wiley & Sons, Chichester, 2nd edition, 1998.
- [109] P. Goodwin and G. Wright. Enhancing Strategy Evaluation in Scenario Planning: a Role for Decision Analysis. *Journal of Management Studies*, 38(1):1–16, 2001.

- [110] M. Goovaerts, F. de Vylder, and J. Haezendonck. *Insurance premiums*. North-Holland Amsterdam, 1984.
- [111] A. Gustafsson, M. Johnson, and I. Roos. The effects of customer satisfaction, relationship commitment dimensions, and triggers on customer retention. *Journal of Marketing*, 69(4):210–218, 2005.
- [112] P. Gyberg and J. Palm. Influencing households’ energy behaviour—how is this done and on what premises? *Energy policy*, 37(7):2807–2813, 2009.
- [113] P. Gyberg and J. Palm. The devil is in the detail: Household electricity saving behaviour and the role of information. *Energy policy*, 38(3):1578–1587, 2009.
- [114] J. Hadar and W. Russell. Rules for Ordering Uncertain Prospects. *American Economic Review*, 59(1):25–34, 1969.
- [115] 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.
- [116] R. Hamalainen. Reversing the Perspective on the Applications of Decision Analysis:(Comment on Keefer et al. 2004). *Decision Analysis*, 1(1):26–31, 2004.
- [117] T. Hanne. Meta decision problems in multiple criteria decision making. In T. Gal, T. Stewart, and T. Hanne, editors, *Multicriteria Decision Making: Advances in MCDM Models, Algorithms, Theory, and Applications*, chapter 6. Kluwer Academic Publishers, Dordrecht, 1999.
- [118] D. Haralambopoulos and H. Polatidis. Renewable energy projects: structuring a multi-criteria group decision-making framework. *Renewable energy*, 28(6):961–973, 2003.
- [119] C. Harries. Correspondence to what? Coherence to what? What is good scenario-based decision making? *Technological Forecasting and Social Change*, 70(8):797–817, 2003.
- [120] Y. He, Q. Wang, and D. Zhou. Extension of the expected value method for multiple attribute decision making with fuzzy data. *Knowledge-based Systems*, 22:63–66, 2009.
- [121] R. Hertwig and G. Gigerenzer. The Conjunction Fallacy Revisited: How Intelligent Inferences Look Like Reasoning Errors. *Journal of Behavioral Decision Making*, 12:275–305, 1999.

- [122] G. Hodgkinson and G. Wright. Confronting strategic inertia in a top management team: learning from failure. *Organization Studies*, 23(6):949, 2002.
- [123] P. Hofstatter. Über die Schätzung von Gruppeneigenschaften. *Zeitschrift für Psychologie*, 145(February, 1939):1–44, 1939.
- [124] J. Hokkanen, R. Lahdelma, K. Miettinen, and P. Salminen. Determining the implementation order of a general plan by using a multicriteria method. *Journal of Multi-Criteria Decision Analysis*, 7(5):273–284, 1998.
- [125] J. Hokkanen, R. Lahdelma, and P. Salminen. A multiple criteria decision model for analyzing and choosing among different development patterns for the Helsinki cargo harbor. *Socio-Economic Planning Sciences*, 33(1):1–23, 1999.
- [126] C. Huang, D. Kira, and I. Vertinsky. Stochastic dominance rules for multiattribute utility functions. *Review of Economic Studies*, 41:611–615, 1978.
- [127] V. Huynh, Y. Nakamori, M. Ryoike, and T. Ho. Decision making under uncertainty with fuzzy targets. *Fuzzy Optimization and Decision Making*, 6(3):255–278, 2007.
- [128] C. Hwang and K. Yoon. *Multiple Attribute Decision Making: Methods and Applications*. Springer Verlag, 1981.
- [129] G. Islei, G. Lockett, and P. Naudé. Judgemental modelling as an aid to scenario planning and analysis. *Omega*, 27(1):61–73, 1999.
- [130] 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.
- [131] S. Jarvenpaa. Graphic displays in decision making: the visual salience effect. *Journal of Behavioral Decision Making*, 3(4):247–262, 1990.
- [132] R. Jensen. An alternative scaling method for priorities in hierarchical structures. *Journal of mathematical psychology(Print)*, 28(3):317–332, 1984.
- [133] J. Jia and J. Dyer. A standard measure of risk and risk-value models. *Management Science*, 42(12):1691–1705, 1996.
- [134] J. Johnson and A. Bruce. Calibration of subjective probability judgments in a naturalistic setting. *Organizational Behavior and Human Decision Processes*, 85(2):265–290, 2001.
- [135] E. Joyce and G. Biddle. Anchoring and adjustment in probabilistic inference in auditing. *Journal of Accounting Research*, pages 120–145, 1981.

- [136] C. Julian-Reynier, M. Welkenhuysen, L. Hagoel, M. Decruyenaere, and P. Hopwood. Risk communication strategies: state of the art and effectiveness in the context of cancer genetic services. *European Journal of Human Genetics*, 11(10):725–736, 2003.
- [137] F. Juster. Consumer buying intentions and purchase probability: an experiment in survey design. *Journal of the American Statistical Association*, 61(315):658–696, 1966.
- [138] D. Kahneman, P. Slovic, and A. Tversky. *Judgment under uncertainty: Heuristics and biases*. Cambridge University Press, 1982.
- [139] D. Kahneman and A. Tversky. Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–291, 1979.
- [140] T. Kalu. Capital budgeting under uncertainty: An extended goal programming approach. *International Journal of Production Economics*, 58:235–251, 1999.
- [141] J. Kangas, J. Hokkanen, A. Kangas, R. Lahdelma, and P. Salminen. Applying stochastic multicriteria acceptability analysis to forest ecosystem management with both cardinal and ordinal criteria. *Forest Science*, 49(6):928–937, 2003.
- [142] C. Kaplan and H. Simon. In search of insight. *Cognitive Psychology*, 22(3):374–419, 1990.
- [143] A. Kaufmann and M. Gupta. *Fuzzy Mathematical Models in Engineering and Management Science*. Elsevier, 1988.
- [144] D. Keifer and S. Bodily. Three-point approximations for continuous random variables. *Management Science*, 29(5):595–609, 1983.
- [145] R. Keeney. *Value-Focussed Thinking*. Harvard University Press, Cambridge, 1992.
- [146] R. Keeney and H. Raiffa. *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*. John Wiley & Sons, New York, 1976.
- [147] R. Keeney and H. Raiffa. *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*. Cambridge University Press, Cambridge, 1993.
- [148] T. Keiningham, T. Perkins-Munn, and H. Evans. The impact of customer satisfaction on share-of-wallet in a business-to-business environment. *Journal of Service Research*, 6(1):37, 2003.
- [149] 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.

- [150] C. Kirkwood. Estimating the impact of uncertainty on deterministic multiattribute evaluation. *Management Science*, 38(6):819–826, 1992.
- [151] D. Kleinmuntz. Decomposition and the control of error in decision analytic models. In H. Einhorn and R. Hogarth, editors, *Insights in decision making: A tribute to Hillel J. Einhorn*, pages 107–126. University of Chicago Press, 1990.
- [152] J. Koehler. The base rate fallacy reconsidered: Descriptive, normative, and methodological challenges. *Behavioral and Brain Sciences*, 19(1):1–16, 1996.
- [153] 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.
- [154] P. Kotler and A. Andreasen. Strategic marketing for nonprofit organizations. *Englewood Cliffs, NJ*, 1996.
- [155] K. Kowalski, S. Stagl, R. Madlener, and I. Omann. Sustainable energy futures: methodological challenges in combining scenarios and participatory multi-criteria analysis. *European Journal of Operational Research*, doi: 10.1016/j.ejor.2007.12.049, 2008.
- [156] M. Kwiesielewicz. The logarithmic least squares and the generalized pseudoinverse in estimating ratios. *European Journal of Operational Research*, 93(3):611–619, 1996.
- [157] M. Kynn. The heuristics and biases bias in expert elicitation. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 171(1):239–264, 2008.
- [158] P. LaBarbera and D. Mazursky. A longitudinal assessment of consumer satisfaction/dissatisfaction: the dynamic aspect of the cognitive process. *Journal of Marketing Research*, 20(4):393–404, 1983.
- [159] R. Lahdelma, J. Hokkanen, and P. Salminen. SMAA – Stochastic multiobjective acceptability analysis. *European Journal of Operational Research*, 106:137–143, 1998.
- [160] R. Lahdelma, K. Miettinen, and P. Salminen. Ordinal criteria in stochastic multi-criteria acceptability analysis (SMAA). *European Journal of Operational Research*, 147(1):117–127, 2003.
- [161] R. Lahdelma, K. Miettinen, and P. Salminen. Reference point approach for multiple decision makers. *European Journal of Operational Research*, 164(3):785–791, 2005.
- [162] R. Lahdelma and P. Salminen. SMAA-2: stochastic multi-criteria acceptability analysis for group decision making. *Operations Research*, 49(3):444–454, 2001.



- [163] R. Lahdelma and P. Salminen. Stochastic multicriteria acceptability analysis using the data envelopment model. *European Journal of Operational Research*, 170(1):241–252, 2006.
- [164] R. Lahdelma and P. Salminen. Prospect theory and stochastic multicriteria acceptability analysis (SMAA). *Omega*, doi: 10.1016/j.omega.2008.09.001, 2008.
- [165] R. Lahdelma, P. Salminen, and J. Hokkanen. Locating a waste treatment facility by using stochastic multicriteria acceptability analysis with ordinal criteria. *European Journal of Operational Research*, 142(2):345–356, 2002.
- [166] R. Lathrop. Perceived variability. *Journal of Experimental Psychology*, 73(4):498, 1967.
- [167] J. Lee and H. Lee-Kwang. Comparison of fuzzy values on a continuous domain. *Fuzzy Sets and Systems*, 118(3):419–428, 2001.
- [168] H. Lee-Kwang and J. Lee. A method for ranking fuzzy numbers and its application to decision-making. *Fuzzy Systems, IEEE Transactions on*, 7(6):677–685, 1999.
- [169] L. Leung and D. Cao. On consistency and ranking of alternatives in fuzzy AHP. *European Journal of Operational Research*, 124(1):102–113, 2000.
- [170] R. Levary and K. Wan. A simulation approach for handling uncertainty in the analytic hierarchy process. *European Journal of Operational Research*, 106(1):116–122, 1998.
- [171] R. Levary and K. Wan. An analytic hierarchy process based simulation model for entry mode decision regarding foreign direct investment. *Omega*, 27(6):661–677, 1999.
- [172] D. Li. Fuzzy multiattribute decision-making models and methods with incomplete preference information. *Fuzzy Sets and Systems*, 106(2):113–119, 1999.
- [173] K. Liang and S. Zeger. Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1):13–22, 1986.
- [174] S. Lichtenstein. Calibration of probabilities: The state of the art to 1980. In *Judgment under uncertainty: Heuristics and biases* [138], pages 306–334.
- [175] S. Lichtenstein and J. Newman. Empirical scaling of common verbal phrases associated with numerical probabilities. *Psychonomic Science*, 9(10):563–564, 1967.
- [176] T. Liou and M. Wang. Ranking fuzzy numbers with integral value. *Fuzzy Sets and Systems*, 50(3):247–255, 1992.

- [177] H. Liu, B. Wu, and M. Liu. Investors' preference order of fuzzy numbers. *Computers & Mathematics with Applications*, 55(11):2623–2630, 2008.
- [178] X. Liu and S. Han. Ranking fuzzy numbers with preference weighting function expectations. *Computers and Mathematics with Applications*, 49(11-12):1731–1753, 2005.
- [179] P. Loba, S. Stewart, R. Klein, and J. Blackburn. Manipulations of the features of standard video lottery terminal (vlt) games: effects in pathological and non-pathological gamblers. *Journal of Gambling Studies*, 17:297–320, 2001.
- [180] A. Lock. Integrating group judgments in subjective forecasts. In Wright and Ayton [287], pages 109–127.
- [181] G. Loomes, C. Starmer, and R. Sugden. Preference Reversal: Information-Processing Effect or Rational Non-Transitive Choice? *Economic Journal*, 99(395):140–151, 1989.
- [182] G. Loomes, C. Starmer, and R. Sugden. Observing violations of transitivity by experimental methods. *Econometrica*, 59(2):425–439, 1991.
- [183] G. Loomes and R. Sugden. Regret theory: An alternative theory of rational choice under uncertainty. *Economic Journal*, 92(368):805–824, 1982.
- [184] 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.
- [185] R. Luce and E. Weber. An axiomatic theory of conjoint, expected risk. *Journal of Mathematical Psychology*, 30:188–205, 1986.
- [186] N. Lurie. Decision making in information-rich environments: The role of information structure. *Journal of Consumer Research*, 30(4):473–486, 2004.
- [187] J. Martel and G. d'Avignon. Projects ordering with multicriteria analysis. *European Journal of Operational Research*, 10:56–69, 1982.
- [188] 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.
- [189] B. Matarazzo and G. Munda. New approaches for the comparison of LR fuzzy numbers: a theoretical and operational analysis. *Fuzzy Sets and Systems*, 118:407–418, 2001.

- [190] P. McCullagh and J. Nelder. *Generalized linear models*. Chapman Hall, London, 2nd edition, 1989.
- [191] L. Mikhailov. A fuzzy programming method for deriving priorities in the analytic hierarchy process. *The Journal of the Operational Research Society*, 51(3):341–349, 2000.
- [192] L. Mikhailov. Deriving priorities from fuzzy pairwise comparison judgements. *Fuzzy Sets and Systems*, 134(3):365–385, 2003.
- [193] L. Mikhailov and P. Tsvetinov. Evaluation of services using a fuzzy analytic hierarchy process. *Applied Soft Computing Journal*, 5(1):23–33, 2004.
- [194] J. Miyamoto and P. Wakker. Multiattribute utility theory without expected utility foundations. *Operations Research*, pages 313–326, 1996.
- [195] D. Montgomery. *Design and analysis of experiments*. Wiley, 1991.
- [196] H. Montgomery. Decision rules and the search for a dominance structure: toward a process model of decision making. In H. Montgomery, P. Humphreys, O. Svenson, and A. Vari, editors, *Analysing and aiding decision processes*. 1983.
- [197] G. Montibeller, H. Gummer, and D. Tumidei. Combining scenario planning and multi-criteria decision analysis in practice. *Journal of Multi-criteria Decision Analysis*, 14:5–20, 2006.
- [198] V. Morwitz, E. Johnson, and D. Schmittlein. Does measuring intent change behavior? *The Journal of Consumer Research*, 20(1):46–61, 1993.
- [199] F. Mosteller and C. Youtz. Quantifying probabilistic expressions. *Statistical Science*, pages 2–12, 1990.
- [200] B. Munier and M. Machina, editors. *Models and experiments in risk and rationality*. Kluwer Academic Publishers, Dordrecht, 1994.
- [201] S. Murakami, H. Maeda, and S. Imamura. Fuzzy decision analysis on the development of centralized regional energy control system. *IFAC Conference on Fuzzy Information*, pages 353–358, 1983.
- [202] A. Murphy and R. Winkler. Credible interval temperature forecasting: some experimental results. *Monthly Weather Review*, 102(11):784–794, 1974.
- [203] I. Newby-Clark, M. Ross, R. Buehler, D. Koehler, and D. Griffin. People focus on optimistic scenarios and disregard pessimistic scenarios while predicting task completion times. *Journal of Experimental Psychology: Applied*, 6(3):171–181, 2000.

- [204] A. Newell and H. Simon. *Human problem solving*. Prentice-Hall Englewood Cliffs, NJ, 1972.
- [205] M. Nowak. Preference and veto thresholds in multicriteria analysis based on stochastic dominance. *European Journal of Operational Research*, 158(2):339–350, 2004.
- [206] P. Odom, R. Shannon, and B. Buckles. Multi-goal subset selection problems under uncertainty. *AIIE Transactions*, 11(1):61–69, 1979.
- [207] A. O’Hagan and J. Oakley. Probability is perfect, but we can’t elicit it perfectly. *Reliability Engineering and system safety*, 85(1-3):239–248, 2004.
- [208] R. Oliver. A cognitive model of the antecedents and consequences of satisfaction decisions. *Journal of Marketing Research*, 17(4):460–469, 1980.
- [209] M. Omodei and A. Wearing. Decision making in complex dynamic settings: A theoretical model incorporating motivation, intention, affect, and cognitive performance. *Sprache & Kognition*, 14(2):75–90, 1995.
- [210] L. Papke and J. Wooldridge. Econometric methods for fractional response variables with an application to 401 (k) plan participation rates. *Journal of Applied Econometrics*, 11(6):619–632, 1996.
- [211] G. Parnell, J. Jackson, R. Burk, L. Lehmkuhl, and J. Engelbrecht. R & D concept decision analysis: Using alternate futures for sensitivity analysis. *Journal of Multi-Criteria Decision Analysis*, 8(3):119–127, 1999.
- [212] J. Payne. It is whether you win or lose: the importance of the overall probabilities of winning or losing in risky choice. *Journal of Risk and Uncertainty*, 30(1):5–19, 2005.
- [213] J. Payne, J. Bettman, and E. Johnson. Behavioral decision research: A constructive processing perspective. *Annual review of psychology*, 43(1):87–131, 1992.
- [214] J. Payne, E. Johnson, and J. Bettman. *The adaptive decision maker*. Cambridge University Press, 1993.
- [215] T. Perkins-Munn, L. Aksoy, T. Keiningham, and D. Estrin. Actual purchase as a proxy for share of wallet. *Journal of Service Research*, 7(3):245, 2005.
- [216] C. Peterson and A. Miller. Mode, median and mean as optimal strategies. *Journal of Experimental Psychology*, 68:363–367, 1964.
- [217] C. Peterson, K. Snapper, and A. Murphy. Credible interval temperature forecasts. *Bulletin of the American Meteorological Society*, 53(10):966–970, 1972.

- [218] J. Quiggin. A theory of anticipated utility. *Journal of Economic Behavior and Organization*, 3(4):323–343, 1982.
- [219] J. Quiggin and R. Chambers. The state-contingent approach to production under uncertainty\*. *Australian Journal of Agricultural and Resource Economics*, 50(2):153–169, 2006.
- [220] P. Raj and D. Kumar. Ranking multi-criterion river basin planning alternatives using fuzzy numbers. *Fuzzy sets and systems*, 100(1-3):89–99, 1998.
- [221] P. Raj and D. Kumar. Ranking alternatives with fuzzy weights using maximizing set and minimizing set. *Fuzzy Sets and Systems*, 105(3):365–375, 1999.
- [222] B. Roy. *Methodologie Multicritere d’Aide à la Decision*. Economica, Paris, 1985.
- [223] B. Roy. Decision-aid and decision-making. *European Journal of Operational Research*, 45(2-3):324–331, 1990.
- [224] R. Rust, K. Lemon, and V. Zeithaml. Return on marketing: using customer equity to focus marketing strategy. *Journal of Marketing*, 68(1):109–27, 2004.
- [225] T. Saaty. *The Analytic Hierarchy Process*. McGraw-Hill, New York, 1980.
- [226] T. Saaty and L. Vargas. Comparison of eigenvalue, logarithmic least squares and least squares methods in estimating ratios. *Mathematical Modelling*, 5(5):309–324, 1984.
- [227] A. Salo and R. Hamalainen. Preference ratios in multiattribute evaluation (PRIME) – elicitation and decision procedures under incomplete information. *IEEE Transactions on Systems, Man, and Cybernetics, Part A*, 31(6):533–545, 2001.
- [228] R. Sarin and M. Weber. Risk-value models. *European Journal of Operational Research*, 70:135–149, 1993.
- [229] P. Schoemaker. When and how to use scenario planning: a heuristic approach with illustration. *Journal of forecasting*, 10(6):549–564, 1991.
- [230] P. Schoemaker. Scenario Planning: A Tool for Strategic Thinking. *Sloan Management Review*, 36:25–25, 1995.
- [231] K. Seiders, G. Voss, D. Grewal, and A. Godfrey. Do satisfied customers buy more? Examining moderating influences in a retailing context. *Journal of Marketing*, 69(4):26–43, 2005.
- [232] J. Shanteau. Competence in experts: The role of task characteristics. *Organizational Behavior and Human Decision Processes*, 53:252–266, 1992.

- [233] H. Simon. *The sciences of the artificial*. MIT Press, 3 edition, 1996.
- [234] V. Singh, K. Hansen, and S. Gupta. Modeling preferences for common attributes in multi-category brand choice. *Journal of Marketing Research*, 42(2):195–209, 2005.
- [235] S. Sloman, D. Over, L. Slovak, and J. Stibel. Frequency illusions and other fallacies. *Organizational Behavior and Human Decision Processes*, 91(2):296–309, 2003.
- [236] J. Smith and T. Kida. Heuristics and biases: Expertise and task realism in auditing. *Psychological Bulletin*, 109(3):472–489, 1991.
- [237] J. Spencer. Estimating averages. *Ergonomics*, 4(4):317–328, 1961.
- [238] C. Spetzler and C.-A. Staël Von Holstein. Probability encoding in decision analysis. *Management Science*, pages 340–358, 1975.
- [239] 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.
- [240] C. Starmer and R. Sugden. Violations of the independence axiom in common ratio problems: an experimental test of some competing hypotheses. *Annals of Operations Research*, 19(1-4):79–102, 1989.
- [241] C. Starmer and R. Sugden. Testing Alternative Explanations of Cyclical Choices. *Economica*, 65(259):347–361, 1998.
- [242] T. Stewart. Simplified approaches for multicriteria decision making under uncertainty. *Journal of Multi-criteria Decision Analysis*, 4:246–258, 1995.
- [243] T. Stewart. Robustness of additive value function methods in MCDM. *Journal of Multi-criteria Decision Analysis*, 5:301–309, 1996.
- [244] T. Stewart. Scenario analysis and multicriteria decision making. In J. Climaco, editor, *Multicriteria Analysis*, pages 519–528. Springer-Verlag, Berlin, 1997.
- [245] T. Stewart. Measurements of risk in fisheries management. *ORiON*, 14(1/2):1–15, 1998.
- [246] T. Stewart. Dealing with uncertainties in MCDA. In J. Figueira, S. Greco, and M. Ehrgott, editors, *Multiple Criteria Decision Analysis: State of the Art Surveys*, chapter 11. Springer, 2005.
- [247] W. Stirling. *Satisficing Games and Decision Making: with applications to engineering and computer science*. Cambridge University Press, 2003.

- [248] H. Sun and J. Wu. A new approach for ranking fuzzy numbers based on fuzzy simulation analysis method. *Applied Mathematics and Computation*, 174(1):755–767, 2006.
- [249] K. Teigen and W. Brun. The directionality of verbal probability expressions: Effects on decisions, predictions, and probabilistic reasoning. *Organizational Behavior and Human Decision Processes*, 80(2):155–190, 1999.
- [250] T. Tervonen, J. Figueira, R. Lahdelma, J. Dias, and P. Salminen. A stochastic method for robustness analysis in sorting problems. *European Journal of Operational Research*, 192(1):236–242, 2009.
- [251] T. Tervonen, H. Hakonen, and R. Lahdelma. Elevator planning with stochastic multicriteria acceptability analysis. *Omega*, 36(3):352–362, 2008.
- [252] T. Tervonen and R. Lahdelma. Implementing stochastic multicriteria acceptability analysis. *European Journal of Operational Research*, 178(3):500–513, 2007.
- [253] L. Tran and L. Duckstein. Comparison of fuzzy numbers using a fuzzy distance measure. *Fuzzy Sets and Systems*, 130(3):331–341, 2002.
- [254] E. Triantaphyllou and C. Lin. Development and evaluation of five fuzzy multiattribute decision-making methods. *International Journal of Approximate Reasoning*, 14(4):281–310, 1996.
- [255] S. Tsaur, T. Chang, and C. Yen. The evaluation of airline service quality by fuzzy MCDM. *Tourism Management*, 23(2):107–115, 2002.
- [256] A. Tversky. Discussion. In D. Bell, H. Raiffa, and A. Tversky, editors, *Decision making: Descriptive, normative, and prescriptive interactions*, pages 599–612. Cambridge University Press, Cambridge, 1988.
- [257] A. Tversky and D. Kahneman. Judgment under uncertainty: heuristics and biases. *Science*, 185:1124–1131, 1974.
- [258] A. Tversky and D. Kahneman. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5(4):297–323, 1992.
- [259] M. Uncles, A. Ehrenberg, and K. Hammond. Patterns of buyer behavior: regularities, models, and extensions. *Marketing Science*, 14(3):71–78, 1995.
- [260] K. Van der Heijden. *Scenarios: The Art of Strategic Conversation*. John Wiley & Sons, New York, 1996.

- [261] P. Van Laarhoven and W. Pedrycz. A fuzzy extension of Saaty's priority theory. *Fuzzy Sets and Systems*, 11:199–227, 1983.
- [262] L. Vargas. Reciprocal matrices with random coefficients. *Mathematical Modelling*, 3(1):69–81, 1982.
- [263] P. Verhoef. Understanding the effect of customer relationship management efforts on customer retention and customer share development. *Journal of Marketing*, 67(4):30–45, 2003.
- [264] J. von Neumann and O. Morgenstern. *Theory of Games and Economic Behavior*. New York, 1953.
- [265] D. Von Winterfeldt and W. Edwards. *Decision Analysis and Behavioural Research*. Cambridge University Press, London, 1986.
- [266] P. Wack. Scenarios: shooting the rapids. *Harvard Business Review*, Nov/Dec:139–150, 1985.
- [267] P. Wack. Scenarios: uncharted water ahead. *Harvard Business Review*, Sept/Oct:73–89, 1985.
- [268] P. Wakker and D. Deneffe. Eliciting von Neumann-Morgenstern utilities when probabilities are distorted or unknown. *Management Science*, pages 1131–1150, 1996.
- [269] M. Walker. The seductiveness of poker machines. In *Proceedings of the 13th National Association for Gambling Studies*, pages 12–27. 2004.
- [270] J. Wallenius, J. Dyer, P. Fishburn, R. Steuer, S. Zionts, and K. Deb. Multiple criteria decision making, multiattribute utility theory: recent accomplishments and what lies ahead. *Management Science*, 54(7):1336–1349, 2008.
- [271] T. Wallsten and D. Budescu. Encoding subjective probabilities: A psychological and psychometric review. *Management Science*, pages 151–173, 1983.
- [272] X. Wang and E. Kerre. Reasonable properties for the ordering of fuzzy quantities (I). *Fuzzy Sets and Systems*, 118(3):375–385, 2001.
- [273] X. Wang and E. Kerre. Reasonable properties for the ordering of fuzzy quantities (II). *Fuzzy Sets and Systems*, 118(3):387–405, 2001.
- [274] Y. Wang and K. Chin. An eigenvector method for generating normalized interval and fuzzy weights. *Applied Mathematics and Computation*, 181(2):1257–1275, 2006.



- [275] Y. Wang and T. Elhag. Fuzzy TOPSIS method based on alpha level sets with an application to bridge risk assessment. *Expert Systems With Applications*, 31(2):309–319, 2006.
- [276] Y. Wang, T. Elhag, and Z. Hua. A modified fuzzy logarithmic least squares method for fuzzy analytic hierarchy process. *Fuzzy Sets and Systems*, 157(23):3055–3071, 2006.
- [277] Y. Wang, Y. Luo, and Z. Hua. On the extent analysis method for fuzzy AHP and its applications. *European Journal of Operational Research*, 186(2):735–747, 2008.
- [278] Y. Wang and C. Parkan. Multiple attribute decision making based on fuzzy preference information on alternatives: Ranking and weighting. *Fuzzy sets and systems*, 153(3):331–346, 2005.
- [279] Y. Wang, J. Yang, D. Xu, and K. Chin. On the centroids of fuzzy numbers. *Fuzzy sets and systems*, 157(7):919–926, 2006.
- [280] Y.-J. Wang and H.-S. Lee. The revised method of ranking fuzzy numbers with an area between the centroid and original points. *Computers & Mathematics with Applications*, 55:2033–2042, 2008.
- [281] Z. Wang, Y. Liu, Z. Fan, and B. Feng. Ranking L–R fuzzy number based on deviation degree. *Information Sciences*, doi:10.1016/j.ins.2008.08.017, 2008.
- [282] 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.
- [283] M. Weck, F. Klocke, H. Schell, and E. Rüenauer. Evaluating alternative production cycles using the extended fuzzy AHP method. *European Journal of Operational Research*, 100(2):351–366, 1997.
- [284] G. Whitmore. Third-degree stochastic dominance. *American Economic Review*, 60(3):457–459, 1970.
- [285] E. Wollenberg, D. Edmunds, and L. Buck. Using scenarios to make decisions about the future: anticipatory learning for the adaptive co-management of community forests. *Landscape and Urban Planning*, 47(1-2):65–77, 2000.
- [286] G. Wright. Changes in the realism and distribution of probability assessments as a function of question type. *Acta Psychologica*, 52(1-2):165–174, 1982.
- [287] G. Wright and P. Ayton, editors. *Judgmental forecasting*. Wiley, Chichester, 1987.

- [288] G. Wright and P. Goodwin. Future-focussed thinking: Combining scenario planning with decision analysis. *Journal of Multi-criteria Decision Analysis*, 8:311–321, 1999.
- [289] G. Wu and R. Gonzalez. Curvature of the probability weighting function. *Management Science*, 42(12):1676–1690, 1996.
- [290] G. Wu and R. Gonzalez. Nonlinear decision weights in choice under uncertainty. *Management Science*, 45(1):74–85, 1999.
- [291] R. Xu. Fuzzy least-squares priority method in the analytic hierarchy process. *Fuzzy sets and systems*, 112(3):395–404, 2000.
- [292] R. Xu and X. Zhai. Fuzzy logarithmic least squares ranking method in analytic hierarchy process. *Fuzzy sets and systems*, 77(2):175–190, 1996.
- [293] Z. Xu and J. Chen. Some models for deriving the priority weights from interval fuzzy preference relations. *European Journal of Operational Research*, 184(1):266–280, 2008.
- [294] Z. Xu and Q. Da. A least deviation method to obtain a priority vector of a fuzzy preference relation. *European Journal of Operational Research*, 164(1):206–216, 2005.
- [295] R. Yager, M. Detyniecki, and B. Bouchon-Meunier. A context-dependent method for ordering fuzzy numbers using probabilities. *Information Sciences*, 138(1-4):237–255, 2001.
- [296] J. Yao and K. Wu. Ranking fuzzy numbers based on decomposition principle and signed distance. *Fuzzy Sets and Systems*, 116(2):275–288, 2000.
- [297] C. Yu. A GP-AHP method for solving group decision-making fuzzy AHP problems. *Computers and Operations Research*, 29(14):1969–2001, 2002.
- [298] H. Zank. Cumulative prospect theory for parametric and multiattribute utilities. *Mathematics of Operations Research*, pages 67–81, 2001.
- [299] K. Zaras. Dominances stochastiques pour deux classes de fonctions d'utilité: concaves et convexes. *Recherche Opérationnelle*, 23:57–65, 1989.
- [300] 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.
- [301] K. Zaras and J. Martel. Multiattribute analysis based on stochastic dominance. In Munier and Machina [200], pages 225–248.

- [302] S. Zeger and K. Liang. Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, 42(1):121–130, 1986.
- [303] H.-J. Zimmerman. *Fuzzy Sets, Decision Making, and Expert Systems*. Kluwer Academic Publishers, Boston, 1987.
- [304] H.-J. Zimmerman. An application-oriented view of modeling uncertainty. *European Journal of Operational Research*, 122:190–198, 2000.

University of Cape Town