



Multi-criteria analysis: a manual



Multi-criteria analysis: a manual

Department for Communities and Local Government
Eland House
Bressenden Place
London
SW1E 5DU
Telephone: 020 7944 4400
Website: www.communities.gov.uk

© Crown copyright 2009

Copyright in the typographical arrangement rests with the Crown.

This publication, excluding logos, may be reproduced free of charge in any format or medium for research, private study or for internal circulation within an organisation. This is subject to it being reproduced accurately and not used in a misleading context. The material must be acknowledged as Crown copyright and the title of the publication specified.

Any other use of the contents of this publication would require a copyright licence. Please apply for a Click-Use Licence for core material at www.opsi.gov.uk/click-use/system/online/pLogin.asp, or by writing to the Office of Public Sector Information, Information Policy Team, Kew, Richmond, Surrey TW9 4DU

e-mail: licensing@opsi.gov.uk

If you require this publication in an alternative format please email alternativeformats@communities.gsi.gov.uk

Communities and Local Government Publications
PO Box 236
Wetherby
West Yorkshire
LS23 7NB
Tel: 0300 123 1124
Fax: 0300 123 1125
Email: communities@capita.co.uk
Online via the Communities and Local Government website: www.communities.gov.uk

January 2009

Product Code: 08ACST05703

ISBN: 978-1-4098-1023-0

Contents

	Preface	5
Chapter 1	The scope and objectives of this manual	6
Chapter 2	Appraisal and evaluation in government	9
	2.1 Introduction	9
	2.2 The decision making process	10
	2.3 Identifying objectives	10
	2.4 Identifying options for achieving the objectives	11
	2.5 Identifying the criteria to be used to compare the options	11
	2.6 Analysis of the options	11
	2.7 Making choices	12
	2.8 Feedback	13
	2.9 Sequences of decisions through time	13
Chapter 3	Monetary-based techniques	14
	3.1 Introduction	14
	3.2 Financial analysis	14
	3.3 Cost-effectiveness analysis	15
	3.4 Cost-benefit analysis	15
	3.5 Inclusion of the results of monetary analyses in an MCA framework	17
Chapter 4	An overview of multi-criteria analysis techniques	19
	4.1 Introduction	19
	4.2 Criteria for selecting MCA techniques	20
	4.3 Key features of MCA	20
	4.4 Different types of MCA	23
	4.5 Direct analysis of the performance matrix	24
	4.6 Multi-attribute utility theory	24
	4.7 Linear additive models	25
	4.8 The Analytical Hierarchy Process	26
	4.9 Outranking methods	26
	4.10 Procedures that use qualitative data inputs	27
	4.11 MCA methods based on fuzzy sets	28
	4.12 Other MCA methods	29
	4.13 Application of MCA in government	29

Chapter 5	Decisions without weights: construction of a performance matrix	30
5.1	The stages of a multi-criteria analysis	30
5.2	Step 1: Establishing the decision context	32
5.3	Step 2: Identifying options	32
5.4	Step 3: Identifying criteria and sub-criteria	32
5.5	Step 4 and beyond (without scoring and weighting)	39
5.6	Step 4: Assessing performance levels (with scoring)	41
Chapter 6	Multi-criteria decision analysis	46
6.1	What is MCDA?	46
6.2	Stages in MCDA	49
6.3	Uncertainty, Risk and MCDA	71
Chapter 7	Case studies	73
7.1	The new approach to appraisal for transport	73
7.2	UK applications of MCDA	77
7.3	The NAO case study: evaluation of overseas trade services	78
7.4	The Nirex case study: appraisal of sites for further investigation as potential repositories for radioactive waste	90
7.5	Medium term revenue budget planning	101
7.6	Lessons from the MCDA case studies	109
	Appendix 1 Continuous MCA models (infinitely variable alternatives)	112
	Appendix 2 Non-compensatory methods	115
	Appendix 3 Multi-attribute utility models	118
	Appendix 4 Linear additive models	121
	Appendix 5 The analytical hierarchy process	127
	Appendix 6 Outranking methods	131
	Appendix 7 Outranking methods with qualitative data	138
	Appendix 8 Fuzzy MCA	140
	Appendix 9 Using MCA to explore stakeholder viewpoints	141
	Appendix 10 Probabilities to preferences	146
	Glossary	149
	Bibliography	154
	Software	161

Preface

This manual was commissioned by the Department for the Environment, Transport and the Regions in 2000 and remains, in 2009, the principal current central government guidance on the application of multi-criteria analysis (MCA) techniques. Since 2000 it has become more widely recognised in government that, where quantities can be valued in monetary terms, MCA is not a substitute for cost-benefit analysis, but it may be a complement; and that MCA techniques are diverse in both the kinds of problem that they address (for example prioritisation of programmes as well as single option selection) and in the techniques that they employ, ranging from decision conferencing to less resource intensive processes.

Chapter 1 The scope and objectives of this manual

This manual provides guidance for Government officials and other practitioners on how to undertake and make the best use of multi-criteria analysis (MCA) for the appraisal of options for policy and other decisions, including but not limited to those having implications for the environment. It covers a range of techniques which can be of practical value to public decision makers and are increasingly being used in the UK and in other countries. They are described in this manual as multi-criteria analysis (MCA) techniques.

The document *Policy Appraisal and the Environment* was published by DETR in 1991.¹ In 1997 a report *Experience with the 'Policy Appraisal and the Environment Initiative'* concluded that there was still room for improvement.² The *Policy Appraisal and the Environment: Policy Guidance* leaflet was published by the DETR in April 1998.³ In the same year a scoping study was commissioned on techniques and published guidance on environment appraisal.⁴ One suggestion for future action was the production of general guidance on multi-criteria analysis. This increasingly popular set of techniques typically combines a range of project, policy or programme option impacts into a single framework for easier assimilation by decision makers. The present manual implements this recommendation.

The manual is about techniques which do not necessarily rely on monetary valuations. It therefore complements guidance on those techniques which primarily use monetary valuations, namely financial analysis, cost effectiveness analysis (CEA), and cost-benefit analysis (CBA). These monetary techniques have been extensively used in UK government circles, and are the subject of a number of guides and manuals.

Most officials are familiar with monetary techniques, but less so with the MCA techniques described in this manual. Nonetheless highway investments, for example, have for many years been appraised using procedures that take account both of impacts measured in monetary units, such as construction costs, time savings and reductions in accident costs, and of social and environmental impacts that may be quantified but not valued (such as the number of houses suffering specified increases in noise) or assessed only in qualitative terms (such as impacts on landscape). In 1998 DETR developed an MCA approach in the form of the New Approach to Appraisal for transport projects. This provides an improved way of presenting monetised and non-monetised impacts of transport projects to decision makers. Other MCA procedures, including 'scoring and weighting' are also already used

¹ Department of the Environment (1991) *Policy Appraisal and the Environment* HMSO, London.

² DETR (1997), *Experience with the 'Policy Appraisal and the Environment' Initiative* KPMG for DETR, London.

³ DETR (1998a) *Policy Appraisal and the Environment: Policy Guidance* DETR, London.

⁴ DETR (1998b) *Review of Technical Guidance on Environmental Appraisal* DETR, London, reporting work produced by EFTEC.

from time to time elsewhere in government. For example, the Environment Agency have been developing the techniques to appraise improvements in water quality, while scoring and weighting systems have been used to assess Structural Fund issues.⁵

This manual provides practical guidance on the application of these techniques, in non-technical language. It is designed to help non-specialist staff to gain an overview of the advantages offered by MCA and what its requirements may be in terms of resources for undertaking appraisals. The manual also has more detailed appendices on various MCA methodologies which will be more accessible to economists and other analytical specialists.

One source of confusion to those new to the field is the variety of different techniques, often with rather similar sounding titles, that appear to be available, such as multi-criteria decision analysis, multi-attribute utility theory, the analytic hierarchy process, and fuzzy set theory. The manual outlines the relationships between the different techniques and indicates the ones which can yield the most fruitful applications, in contrast to those which may be of theoretical interest but little practical value.

Chapters 2–4 of the manual provide a broad overview of the techniques for non-specialists. Chapter 2 takes an overview of appraisal and evaluation in government. Chapter 3 briefly reviews the monetary-based techniques. Chapter 4 provides a non-technical overview of the main MCA techniques that are not centred on monetary valuation. It presents criteria for choosing between them, and introduces the techniques which will be discussed in detail in the rest of the manual.

Chapters 5–7 provide more detailed guidance for specialists interested in applying the techniques to their own particular decision-making problems. Chapter 5 sets out the stages involved in carrying out a multi-criteria analysis, up to and including the construction of a performance matrix, which sets out how each of the options being appraised performs on each of the criteria that form part of the analysis. Chapter 6 shows how to carry out a full multi-criteria decision analysis (MCDA) involving scoring of each option on each criterion, and then combining the scores by means of a system of weights to yield an overall ranking for each option. As well as providing an overview of the process as a whole, in a way consistent with that set out in Chapter 5, the chapter also takes the reader through each stage, step-by-step, and illustrates the process with an example of a simple household decision making process. Chapter 7 describes a number of case studies. These include DETR's New Approach to Appraisal, which uses a form of performance matrix known as the Assessment Summary Table, and three cases of MCDA. The MCDA cases are an evaluation of different services to exporters which formed part of a National Audit Office investigation, an appraisal of different sites for nuclear waste disposal, and a local authority's review of objectives in the context of planning expenditure.

⁵ See MAFF, Economics (Resource Use) Division (1996) *Scoring, Weighting and Ranking (SWR) Systems to Assist Structural Fund Project Selection*; and European Union (1995) *Applying the Multi-Criteria Method to the Evaluation of Structural Programmes*, Means Handbook 4.

In order to aid the reader's understanding, technical terms are explained in the glossary. Italics in the main text indicate the first appearance of each of these technical terms.

A guide to further reading provides advice on main sources of additional information in print.

Finally, a guide to software describes the main software tools available in the spring of 2000 to apply the techniques in practice.

To get an idea of the application of a form of MCA, you could now read through the example of the use of multi-criteria decision analysis in the boxes in Chapter 6.

Chapter 2 Appraisal and evaluation in government

2.1 Introduction

In practice the most common form of analysis in government is *cost-effectiveness analysis* (CEA), where the costs⁶ of alternative ways of providing similar kinds of output are compared. Any differences in output are compared subjectively with the differences in costs. The Treasury 'Green Book' on *Appraisal and Evaluation in Central Government* is most frequently applied to CEA.⁷

Less common, although widely used in transport and health and safety, is *cost benefit analysis* (CBA),⁸ in which some important non-marketed outputs are explicitly valued in money terms.

Both CEA and CBA are analytical ways of comparing different forms of input or output, in these cases by giving them money values, and might themselves be regarded as examples of multi-criteria analysis. However this manual is concerned with techniques for comparing impacts in ways which do not involve giving all of them explicit monetary values, although they may include data from cost-effectiveness or cost-benefit analyses.

The techniques described in this manual are in many respects an 'alternative' to defining monetary values for all the major costs and benefits when this is impractical. However MCA must not be seen as a short cut, nor as an easier technique for inexperienced people to use. The use of these techniques is in important ways more demanding of experience and good training than the use of CEA, which generally uses market values, or of CBA, which also uses valuations of non-marketed quantities based on analysis which has already been completed elsewhere.

Criteria and Attributes

The words criterion and attribute are often used synonymously in the literature on MCA, which is indeed sometimes referred to as multi-attribute analysis. Attribute is also sometimes used to refer to a measurable criterion. In this manual we use the word criterion rather than attribute.

⁶ These costs may include both financial costs, and wider opportunity costs.

⁷ HM Treasury (1997) *Appraisal and Evaluation in Central Government: Treasury Guidance*. "The Green Book" The Stationery Office, London. The Green Book includes an annex on "Costs and benefits not easily valued", which briefly discusses both CBA and MCA approaches.

⁸ Sometimes prefixed as **social** cost-benefit analysis, since firms may use the term cost-benefit analysis to refer to a process of comparing the financial costs and benefits of particular actions. However, in this manual we use the term cost-benefit analysis to refer to the technique which tries to measure all the social costs and benefits of an action, and reserve financial analysis for comparisons of the financial consequences of actions to a firm, a government agency, or a government department.

2.2 The decision making process

Decision making about proposals for future action should normally follow the sequence below. For the assessment of how past decisions have worked out in practice some of the headings would be slightly different, but the same principles apply. This manual uses the usual UK central government terminology of describing the analysis of proposed actions as *appraisal* and the retrospective analysis of how actions have worked out in practice as *evaluation*.⁹

The following process might apply to the development of a policy, a programme or a project.

- Identifying objectives
- Identifying options for achieving the objectives
- Identifying the criteria to be used to compare the options
- Analysis of the options
- Making choices, and
- Feedback.

Each is considered in turn below.

2.3 Identifying objectives

Good decisions need clear *objectives*. These should be specific, measurable, agreed, realistic and time-dependent.

It is sometimes useful to classify objectives according to their level. For example, the Treasury Green Book distinguishes between ultimate, intermediate and immediate objectives, but it is particularly useful to distinguish between ultimate and immediate ones.

- Ultimate objectives are usually framed in terms of strategic or higher-level variables, such as the level of economic growth, social cohesion or sustainable development. These objectives may be stated in White Papers, or in Departmental or Agency plans or in annual reports.
- Immediate objectives are those which can be directly linked with the outputs of the policy, programme, or project. Consideration of a proposed option needs to concentrate on those *criteria* which contribute to the immediate, and hence to the ultimate objectives.

The issue of 'whose objectives' should be represented in MCA is discussed in the box in section 2.6.

⁹ This distinction between appraisal and evaluation is a rather specific one, and the terms appraisal and, especially, evaluation are often used with a wide range of meanings outside central government.

2.4 Identifying options for achieving the objectives

Once the objectives are defined, the next stage is to identify options that may contribute to the achievement of these objectives. Options may range from broad policies, such as new environmental priorities for the transport sector, through to the choice of particular lines of routes for roads or the selection of individual projects to improve water quality.

Potentially sensible options then need to be developed in detail. This may range from broad policy design, such as the design of tax policy, through to the more detailed design of individual investment projects. There can be an important feedback to the design stage from all the subsequent stages of appraisal/evaluation.

2.5 Identifying the criteria to be used to compare the options

The next stage is to decide on how to compare different options' contribution to meeting the objectives. This requires the selection of criteria to reflect performance in meeting the objectives. Each criterion must be measurable, in the sense that it must be possible to assess, at least in a qualitative sense, how well a particular option is expected to perform in relation to the criterion.

2.6 Analysis of the options

The next stage in the process is analysis.

Common forms of analysis in government are financial analysis, cost effectiveness analysis and, in some areas, cost benefit analysis, all of which rely wholly or very largely on monetary valuations. Each of these is briefly described in Chapter 3, to provide a context for the MCA techniques with which this manual is concerned.

Where issues of scientific analysis of relevant evidence are concerned, and particularly where there is scientific uncertainty, a range of opinion or potentially significant implications for sensitive areas of public policy, then the principles published in the 'May Guidelines' on scientific advice and policy making should be considered fully.¹⁰

¹⁰ Office of Science and Technology (2000), *Guidelines 2000: Scientific Advice and Policy Making*, Office of Science and Technology, London.

Whose objectives?

Analysis which is carried out in monetary terms does not usually, in practice, present the analysts with problems of choosing between the interests of different groups in society. A cost-effectiveness appraisal may be comparing options with slightly different outputs, but these differences will typically be weighted by decision makers at a later stage. The valuations used in cost-benefit analysis will include distributional judgements, but these will have been determined at an earlier stage.

The techniques described in this manual are more likely to require the analyst to provide distributional judgements. For example both the scoring and weighting stages build in judgements about whose preferences the scores and weights represent.

Such judgements will vary from case to case, depending upon, for example, institutional mindsets and the known preferences of the government of the day. The Treasury Green Book states that analysis within government is concerned with effects “on the national interest”. However different institutions might interpret this in different ways, to reflect for example the views of experts, Ministers, senior officials, public opinion, or those directly affected by the decision. A broadly satisfactory criterion which appears to underlie many CBA valuations is that they should reflect the **informed** preferences of people as a whole, to the extent that these preferences can be measured and averaged. This argues in favour of ensuring that the objectives included in any MCA analysis are sufficiently wide to encompass the main concerns of people as a whole. But after analysis, there will always be further strategic or pragmatic issues to which those responsible for final decisions must also give weight. As a general guide, this manual can go no further than identifying the issue of whose objectives should be represented as an issue which the analyst should recognise and address explicitly rather than implicitly.

2.7 Making choices

The final stage of the decision making process is the actual choice of option. This needs to be seen as a separate stage because none of the techniques available, whether they be financial analysis, cost-benefit analysis, or the different forms of multi-criteria analysis, can incorporate into the formal analysis every judgement, for example about future changes in the state of the world, or income distribution, or political impact, which the ultimate decision needs to take into account. The final decision may sometimes be taken by officials and sometimes by Ministers, depending on its political content.

Even at this stage it may be decided that a further option or options should be considered and the analysis revisited.

2.8 Feedback

Good decision making requires a continuous reassessment of choices made in the past. Individual decision makers may learn from their own mistakes, but it is important that lessons be learned in a more formal and systematic way, and communicated to others, so that they can inform future decisions.

2.9 Sequences of decisions through time

An area of decision making closely related to the topics covered in this manual involves the use of *decision trees* to help identify good strategies for planning a response to a set of interdependent decisions sequenced through time. The actual outcome of each of the individual decisions at each stage is not known with certainty. Appropriate analysis of the tree allows the decision maker to develop, from the outset of the decision process, a contingent decision strategy.

Decision trees have as their prime focus the question of uncertainty about the outcomes of decisions and, in general, pay little attention to the way in which individual decision outcomes are appraised. Rather, they reflect relatively simple appraisal guidelines, such as straightforward maximisation of profit or minimisation of cost. They are often, though not exclusively, applied to problems which are analytically well defined and well supported by technical data. In contrast, this manual devotes substantial attention to appraisal of, and choice between, outcomes often in circumstances where the basis for appraising them is not immediately clear. Many of the most challenging public policy decisions are of this latter type and the manual focuses on this area.

Risk and uncertainty are, however, important issues for most types of decision making and therefore also feature in this manual. Good introductions to decision trees and the types of problem to which they can usefully be applied can be found in Golub (1997) and Targett (1996). For problems that merit it, it is also possible to combine decision tree modelling with the appraisal principles set out in this manual, and in this way develop contingent decision strategies based on multi-criteria assessment.¹¹

¹¹ References: Golub, A.L., *Decision Analysis: an Integrated Approach*, John Wiley, 1997. Targett, D., *Analytical Decision Making*, Pitman Publishing, 1996, chapter 3.

Chapter 3 Monetary-based techniques

3.1 Introduction

This chapter outlines the various decision-support techniques which are based primarily on monetary valuation of the impacts of options. These techniques are:

- **Financial analysis.** An assessment of the impact of an option on the decision-making organisation's own financial costs and revenues.
- **Cost-effectiveness analysis.** An assessment of the costs of alternative options which all achieve the same objective. The costs need not be restricted to purely financial ones.
- **Cost-benefit analysis.** An assessment of all the costs and benefits of alternative options.

The rest of the chapter is structured as follows. Section 3.2 considers financial analysis, section 3.3 considers cost-effectiveness analysis, and section 3.4 considers cost-benefit analysis, including its limitations. This leads on to the discussion of MCA techniques in the rest of this manual. However, before doing so, section 3.5 indicates how financial analysis, cost-effectiveness analysis, or cost-benefit analysis may themselves form an input into MCA. The discussion of cost-benefit analysis leads into the discussion of the DETR's current appraisal procedures for road investment, which we examine in greater detail in Chapter 7.

3.2 Financial analysis

For any significant proposal the relevant department or agency will generally carry out a financial analysis of the impacts on its own budget, and on public expenditure. This includes Regulatory Impact Assessments (RIAs), which are required whenever proposals for regulations or legislation are published, and are intended to help reach decisions about whether the proposed measures are needed.

If the impacts are spread over future years, the net impacts in each year need to be *discounted* to a present value, and this applies equally to cost effectiveness and cost-benefit analysis.

3.3 Cost-effectiveness analysis

Where there are alternative options to achieve a specific objective, but where the objective itself cannot be valued, cost-effectiveness analysis can be used to assess the least-cost way of achieving the objective. (Care needs to be taken that differences in output quality, for example because of differences in service quality, are not overlooked.)

As indicated in Chapter 2, the Treasury Green Book is most frequently concerned with cost-effectiveness analysis.

Cost-effectiveness analysis should certainly include non-cash opportunity costs, such as the use of assets owned by the spending body, which would otherwise be put to some other use. It may also include external costs, if these are relatively straightforward to value in monetary terms – such as taxpayer costs of compliance with changes in tax legislation, or travel costs of out-patients to different hospitals for treatment.

It may possibly include *shadow prices* for some marketed inputs. There are many reasons why market prices do not exactly measure relative impacts on national welfare and shadow prices are sometimes used to reflect these differences between market prices and social opportunity costs. For example different methods of provision may attract different levels of general taxation, which should not be allowed to distort the analysis. These issues are covered in existing central government guidance, which advises that, in a developed economy, market prices are in practice adequate in all but a few special cases.

3.4 Cost-benefit analysis

Cost-benefit analysis seeks to value the expected impacts of an option in monetary terms. These *valuations* are based on a well-developed economic theory of valuation based on willingness-to-pay or to accept. This theory can act as a guide to how valuation should be achieved, and as a referee in disputes about valuation.

The valuations are based on the willingness to pay of the potential gainers for the benefits they will receive as a result of the option, and the willingness of potential losers to accept compensation for the losses they will incur. In principle, a project is desirable if the benefits exceed the losses, suitably discounted over time. This is referred to as the potential compensation principle, since compensation will not normally actually be paid, and so there will be both winners and losers. Since willingness-to-pay and willingness-to-accept will be partly dependent on incomes, there can be a case for weighting gains and losses to take account of income distribution. This is rarely done in practice in the UK because of (1) the lack of consensus on what if any weighting system is appropriate in particular circumstances for people with different incomes and (2) the additional measurement difficulty

of tracing out who ultimately gains and loses from actions once the value of initial benefits and costs have been determined.¹²

Many valuation techniques arise in cost-benefit analysis, and most cost-benefit studies draw on a pool of experience of methods and actual values to complete their analysis.

An option may for example have environmental impacts on noise, local air quality, climate change, water quality, biodiversity, landscape, or townscape. There is an extensive, although by no means always conclusive, literature on the valuation of impacts of this general kind. Two widely used methods are *hedonic price techniques* and *stated preference*. Hedonic price techniques use market valuations such as house prices which may contain information about people's values for other characteristics such as environmental quality, and try to separate out the environmental component using statistical techniques. The use of observed price information for a revealed preference approach contrasts with stated preference methods which seek more direct consumer valuations of environmental effects by asking individuals about their willingness-to-pay, or willingness to accept compensation, for specified changes in environmental quality.

In practice it is hardly ever realistic to value all the costs and benefits of options in monetary terms. Most cost-benefit analyses will incorporate some additional items which it is either not possible to value, or not economic to do so. But where the most important costs and benefits have been valued, the others can be set alongside and included in the decision process.

CBA has great attractions as a tool for guiding public policy:

- it considers the gains and losses to all members of the society on whose behalf the CBA is being undertaken
- it values impacts in terms of a single, familiar measurement scale – money – and can therefore in principle show that implementing an option is worthwhile relative to doing nothing
- the money values used to weight the relative importance of the different impacts are based on people's preferences generally using established methods of measurement.

3.4.1 THE LIMITATIONS OF COST-BENEFIT ANALYSIS

Sometimes CBA is criticised on political or philosophical grounds, to the effect that it is the role of government to apply judgements that are not necessarily a reflection of current preferences in fields such as, for example, environmental degradation. Views on this differ, according to people's views on the role of government. However it is not in practice a major obstacle.¹³

¹² Initial and ultimate benefits may differ for a number of reasons. Where an investment project reduces the costs of supplying an intermediate input, such as freight transport or electricity for industrial users, the initial benefits of reduced costs to road hauliers or power supply companies will eventually be largely passed on to the consumers of the products and services produced using these inputs. In other circumstances the initial benefits of projects, such as reduced transport costs as the result of a light rail scheme, may be passed on to the owners of properties in the areas served by the system through changes in house prices. Considerations of this kind can be as important in MCA as in CBA.

¹³ In practice it is common for governments to provide some input to CBA valuations where there is a national dimension, beyond the immediate interests of the individual. For example, time savings and accident risks to travellers are generally

Much more serious is the fact that, while procedures such as stated preference or hedonic pricing provide ways to establish money values of some non-marketed impacts, for others it is not immediately practicable. Relevant data may not be available or may be too expensive to collect. It may not be possible to present some impacts in terms where people are able to make reliable trade-offs against money.

In addition, there may be impacts which cannot readily be quantified in a way which could be set against a scale of monetary values. The number of deaths or injuries saved by a safety improvement, or the time saved by a public transport investment, can typically be quantified precisely and valued against a predetermined monetary scale. It may also be possible, by a special study, to derive the value of a specific environmental cost or improvement to those whom it affects. However the effects of a proposed government measure on outputs with diffuse social consequences, such as social cohesion, are often issues on which Ministers wish to apply their own views at the time of decision.

CBA is also sometimes criticised for the limitation that it does not generally take account of the interactions between different impacts. For example, people might feel more strongly negative about a project that imposes both environmental and social costs than would be estimated by adding separate valuations of the two effects.

The techniques presented in this manual do not enjoy all the strengths of CBA. They also share some of the problems of CBA and have important further problems of their own. However, they also have important strengths of their own. In particular they can provide more flexibility than CBA, and are more comprehensive in their coverage.

3.5 Inclusion of the results of monetary analyses in an MCA framework

All cost-benefit studies entail elements which are identified as relevant impacts, but which are not valued. In some circumstances they may be regarded as relatively minor, and so will be listed in the CBA report alongside the overall estimates of those net social benefits which can be valued. They may reinforce the choice ordering implied by the monetary results, or they may not be regarded as sufficient to change this ordering, or sometimes, where the difference between alternatives implied by monetary valuations is small, they may tip the balance.

However, in other circumstances, there may be items for which satisfactory values have not been derived, but which are nevertheless regarded as being of major importance. In these circumstances MCA techniques may be useful.

An example of this type of situation is that of road and other transport infrastructure. Cost-benefit analysis has been used to appraise road investment in the UK since the early 1960s, while there have been extensive valuation studies of the environmental consequences of highway and other transport schemes. However, it is widely recognised that:

- Transport investment schemes do have major environmental consequences which are relevant to public decision making; and
- There are no generally accepted valuations for these environmental effects, in the way that there are for travel time values and values of accidents.

The approach used since the publication of the Leitch Report of the Advisory Committee on Trunk Road Assessment in 1977 has been to incorporate these non-monetary elements into the decision framework by measuring them on numerical scales or by including qualitative descriptions of the effects, in an Environmental Impact Statement. The approach was further formalised in 1998 in DETR's 'New Approach to the Appraisal of Roads Schemes', with its use of Appraisal Summary Tables. This was used in the decisions which Ministers reached about the schemes to be included in the Targeted Programme of Improvements, details of which were published in the Roads Review.¹⁴ This is one of the case studies we discuss in Chapter 7 of this manual.

More generally, whenever some costs and benefits **can** be valued in monetary terms, either by direct observation of prices if appropriate or indirectly using generally accepted techniques, then these data should naturally be used within any wider MCA. As we will show throughout the manual, MCA applications often involve combinations of some criteria which are valued in monetary terms, and others for which monetary valuations do not exist.¹⁵

Difficulties arise when some of the monetary valuations are not regarded as very robust. In the longer term more resources might be used to try to improve their accuracy, but in the immediate term the appraisers might consider either the use of sensitivity analysis to see how much results depends on the particular values used, or whether it might be more appropriate to ignore the (rather mistrusted) monetary values and rely on some more subjective scoring and weighing systems to reflect decision makers' or interest groups' preferences.

¹⁴ See DETR (1998a) *A New Deal for Trunk Roads in England: Understanding the New Approach to Appraisal*, DETR, London.

¹⁵ This principle may be familiar to environmental economists who sometimes use the term Total Economic Value. This distinguishes the various components including an item's actual use value which may be reflected in market prices or may be estimated by hedonic methods, its option value, existence value and intrinsic or other possible values that some people may ascribe to it.

Chapter 4 An overview of multi-criteria analysis techniques

4.1 Introduction

Although all the techniques to be described in this manual would be widely acknowledged as methods of multi-criteria analysis, they cover a wide range of quite distinct approaches (in contrast notably to CBA, which is a more unified body of techniques). Some kinds of MCA do not at present offer much help for practical decision taking, but some can be of considerable value. This manual describes and explains these practical techniques, and indicates the types of application in which they may be used.

All MCA approaches make the options and their contribution to the different criteria explicit, and all require the exercise of judgement. They differ however in how they combine the data. Formal MCA techniques usually provide an explicit relative weighting system¹⁶ for the different criteria.

The main role of the techniques is to deal with the difficulties that human decision-makers have been shown to have in handling large amounts of complex information in a consistent way.

MCA techniques can be used to identify a single most preferred option, to rank options, to short-list a limited number of options for subsequent detailed appraisal, or simply to distinguish acceptable from unacceptable possibilities.

As is clear from a growing literature, there are many MCA techniques and their number is still rising. There are several reasons why this is so:

- there are many different **types of decision** which fit the broad circumstances of MCA
- the **time** available to undertake the analysis may vary
- the amount or nature of **data** available to support the analysis may vary
- the **analytical skills** of those supporting the decision may vary, and
- the **administrative culture and requirements** of organisations vary.

This chapter gives a broad overview of the full range of MCA techniques currently available. However, it is neither necessary nor desirable to explore all these techniques in detail. Some are oriented towards issues which public sector decision makers are unlikely to encounter; some are complex and untested in practice; others lack sound theoretical foundations.

¹⁶ Nearly all decisions imply some form of weighting system, though perhaps implicit, and not necessarily consistent. A limited exception, dominance, is described in Section 5.5.2.1.

4.2 Criteria for selecting MCA techniques

Criteria used in this manual for the selection of techniques¹⁷ are:

- internal consistency and logical soundness
- transparency
- ease of use
- data requirements not inconsistent with the importance of the issue being considered
- realistic time and manpower resource requirements for the analysis process
- ability to provide an audit trail, and
- software availability, where needed.

4.3 Key features of MCA

Multi-criteria analysis establishes preferences between options by reference to an explicit set of objectives that the decision making body has identified, and for which it has established measurable criteria to assess the extent to which the objectives have been achieved. In simple circumstances, the process of identifying objectives and criteria may alone provide enough information for decision-makers. However, where a level of detail broadly akin to CBA is required, MCA offers a number of ways of aggregating the data on individual criteria to provide indicators of the overall performance of options.

A key feature of MCA is its emphasis on the judgement of the decision making team, in establishing objectives and criteria, estimating relative importance weights and, to some extent, in judging the contribution of each option to each performance criterion. The subjectivity that pervades this can be a matter of concern. Its foundation, in principle, is the decision makers' own choices of objectives, criteria, weights and assessments of achieving the objectives, although 'objective' data such as observed prices can also be included. MCA, however, can bring a degree of structure, analysis and openness to classes of decision that lie beyond the practical reach of CBA.

One limitation of MCA is that it cannot show that an action adds more to welfare than it detracts. Unlike CBA, there is no explicit rationale or necessity for a Pareto Improvement rule that benefits should exceed costs. Thus in MCA, as is also the case with cost effectiveness analysis, the 'best' option can be inconsistent with improving welfare, so doing nothing could in principle be preferable.

¹⁷ One could regard the choice between techniques as an example of a decision that could be made with the assistance of MCA. There are a number of criteria, and the performance of each technique could be scored against each criterion, and the different criteria weighted, to give an overall preference ordering between different techniques. Except... to do this one would first have to decide which technique to use to assess the techniques!

4.3.1 ADVANTAGES OF MCA OVER INFORMAL JUDGEMENT

MCA has many advantages over informal judgement unsupported by analysis:

- it is open and explicit
- the choice of objectives and criteria that any decision making group may make are open to analysis and to change if they are felt to be inappropriate
- scores and weights, when used, are also explicit and are developed according to established techniques. They can also be cross-referenced to other sources of information on relative values, and amended if necessary
- performance measurement can be sub-contracted to experts, so need not necessarily be left in the hands of the decision making body itself
- it can provide an important means of communication, within the decision making body and sometimes, later, between that body and the wider community, and
- scores and weights are used, it provides an audit trail.

4.3.2 THE PERFORMANCE MATRIX

A standard feature of multi-criteria analysis is a *performance matrix*, or consequence table, in which each row describes an option and each column describes the performance of the options against each criterion. The individual performance assessments are often numerical, but may also be expressed as 'bullet point' scores, or colour coding. Table 4.1 shows a simple example, which we will consider in our detailed introduction to MCDA in Chapter 6. The table, based on an analysis in *Which?* magazine,¹⁸ shows the performance of a number of different toasters in regard to a set of criteria thought to be relevant in a household's choice between different models. These criteria are price, presence of reheat setting, warming rack, adjustable slot width, evenness of toasting, and number of drawbacks. As can be seen, some of these criteria are measured in cardinal numbers (price, number of drawbacks), some in binary terms (a tick indicates presence of a particular feature), and one in qualitative terms (evenness of toasting).

In a basic form of MCA this performance matrix may be the final product of the analysis. The decision makers are then left with the task of assessing the extent to which their objectives are met by the entries in the matrix. Such intuitive processing of the data can be speedy and effective, but it may also lead to the use of unjustified assumptions, causing incorrect ranking of options. This issue is discussed further in Section 5.5.2.3.

In analytically more sophisticated MCA techniques the information in the basic matrix is usually converted into consistent numerical values. In Chapter 6 we show how this can be done using the toaster example.

¹⁸ *Which?* (November 1995), published by Consumers' Association, 2 Marylebone Road, London NW1 4DF.

Table 4.1 Performance matrix

Options	Price	Reheat setting	Warming rack	Adjustable slot width	Evenness of toasting	Number of drawbacks
Boots 2-slice	£18				☆	3
Kenwood TT350	£27	✓	✓	✓	☆	3
Marks & Spencer 2235	£25	✓	✓		★	3
Morphy Richards Coolstyle	£22				☆	2
Philips HD4807	£22	✓			★	2
Kenwood TT825	£30				☆	2
Tefal Thick'n'Thin 8780	£20	✓		✓	★	5

A tick indicates the presence of a feature. Evenness of toasting is shown in *Which?* on a five-point scale, with a solid star representing the best toaster, and an open star the next best. The family eliminated from consideration all the toasters that scored less than best or next best.

4.3.3 SCORING AND WEIGHTING

MCA techniques commonly apply numerical analysis to a performance matrix in two stages:

1. Scoring: the expected consequences of each option are assigned a numerical score on a strength of preference scale for each option for each criterion. More preferred options score higher on the scale, and less preferred options score lower. In practice, scales extending from 0 to 100 are often used, where 0 represents a real or hypothetical least preferred option, and 100 is associated with a real or hypothetical most preferred option. All options considered in the MCA would then fall between 0 and 100.
2. Weighting: numerical weights are assigned to define, for each criterion, the relative valuations of a shift between the top and bottom of the chosen scale.

Mathematical routines, which may be written into computer programmes, then combine these two components to give an overall assessment of each option being appraised. This approach therefore requires individuals to provide those inputs that they are best suited to provide, and leaves computers the task of handling detailed information in a way that is consistent with the preferences that have been revealed by these human inputs.

These approaches are often referred to as *compensatory MCA techniques*, since low scores on one criterion may be compensated by high scores on another. The most common way to combine scores on criteria, and relevant weights between criteria, is to calculate a simple weighted average of scores. The discussion of MCA techniques with explicit weights in this manual concentrates on such simple weighted averages.

Use of such weighted averages depends on the assumption of *mutual independence of preferences*. This means that the judged strength of preference for an option on one criterion will be independent of its judged strength of preference on another. Later in the manual, the assumption will be explained in detail, procedures for testing its validity will be provided,

and its role in detecting double-counting of criteria will be explained. Where mutual independence of preferences cannot be established, other MCA procedures are available, although they tend to be more complex to apply.

4.4 Different types of MCA

As indicated, there are many different MCA procedures. This manual concentrates on a small suite of them, selected using the criteria set out in section 4.2 and oriented towards types of decision commonly faced by public sector bodies. Nonetheless, it will be useful to undertake a brief review of the field as a whole as other methods may be encountered from time to time in other applications.

An important initial consideration in the choice of MCA technique is that of the **number of alternatives to be appraised**. Some problems, especially in design and engineering, are concerned with outcomes that are infinitely variable. However, most policy decisions, even at fairly low levels, are usually about choices between discrete options, for example, between alternative investment projects, or between alternative types of tax system. This manual is concerned mainly with techniques for handling choices between a finite number of options. Solving problems involving optimising infinitely variable quantities requires quite different types of procedure, and a brief introduction to these is provided in Appendix 1.

Where the number of options is finite, it does not matter in principle whether this number is small or large. However, it is important to bear in mind that each option that has to be considered has to be appraised to determine how well it performs on each of its criteria. Gathering and processing these data will consume resources, the more so if a large number of criteria has been identified. In choosing whether to implement one of the simpler or one of the more detailed MCA decision support procedures, this is a factor to bear in mind.

In MCA problems with a finite number of options, each of which is assessed in terms of a given number of criteria, the initial frame of reference is essentially the performance matrix described in section 4.3.2. For each option, with respect to each criterion, this performance information needs to be collected. Chapter 5 explains the considerations to be taken into account in constructing and interpreting such matrices.

MCA procedures are distinguished from each other principally in terms of how they process the basic information in the performance matrix. Different circumstances will be better suited to some MCA procedures than others. The rest of this chapter briefly summarises the main features of some of the better known methods and the relationships between them. It does not explore technical detail. In some cases, more detailed background material is given in appendices. For the MCA methods which this manual presents as particularly suited to the public sector, full explanations of the techniques follow in Chapters 5 and 6.

4.5 Direct analysis of the performance matrix

A limited amount of information about options' relative merits can be obtained by direct inspection of the performance matrix. An initial step can be to see if any of the options are dominated by others.

Dominance occurs when one option performs at least as well as another on all criteria and strictly better than the other on at least one criterion. In principle, one option might dominate all others, but in practice this is unlikely. When it does occur, it is helpful to ask if there is some advantage of the dominated option that is not represented by the criteria; this may reveal new criteria that have been overlooked. Dominance is more likely just to enable the decision-making team to eliminate dominated options from further consideration. Section 5.5.2.1 below considers dominance in further detail.

Once any dominance analysis has been concluded, the next stage is for the decision-making team to determine whether trade-offs between different criteria are acceptable, so that good performance on one criterion can in principle compensate for weaker performance on another. Most public decisions admit such trade-offs, but there may be some circumstances, perhaps where ethical issues are central, where trade-offs of this type are not acceptable. If it is not acceptable to consider trade-offs between criteria, then there are a limited number of *non-compensatory MCA* techniques available. Appendix 2 outlines some of these methods, but in general they are not very effective in distinguishing between options in real applications.

Where compensation is acceptable, most MCA methods involve implicit or explicit aggregation of each option's performance across all the criteria to form an overall assessment of each option, on the basis of which the set of options can be compared. The principal difference between the main families of MCA methods is the way in which this aggregation is done. The following sections outline some of the best-known approaches.

4.6 Multi-attribute utility theory

There is no *normative* model of how individuals should make multi-criteria choices that is without critics. The one that comes closest to universal acceptance is based on multi-attribute utility theory and derives from the work of von Neumann and Morgenstern,¹⁹ and of Savage,²⁰ in the 1940s and 1950s.²¹

While this work provided powerful theoretical insights, it does not directly help decision makers in undertaking complex multi-criteria decision tasks.

¹⁹ Von Neumann, J. and Morgenstern, O. (1947) *Theory of Games and Economic Behaviour*, second edition, Princeton University Press, Princeton.

²⁰ Savage, L.J. (1954) *The Foundations of Statistics* Wiley, New York, NY

²¹ In subsequent years some limitations have become apparent in the ability of this line of theory to express the preferences of individuals. However it remains relevant to most areas of government decision making about marginal impacts on community welfare.

The breakthrough in this respect is the work of Keeney and Raiffa, published in 1976.²² They developed a set of procedures, consistent with the earlier normative foundations, which would allow decision makers to evaluate multi-criteria options in practice (Appendix 3).

There are three building blocks for their procedures. First is the performance matrix (section 4.3.2) and the second is procedures to determine whether criteria are independent of each other or not. The third consists of ways of estimating the parameters in a mathematical function which allow the estimation of a single number index, U , to express the decision maker's overall valuation of an option in terms of the value of its performance on each of the separate criteria.

The Keeney and Raiffa approach to decision support has been applied to many real decisions, in both the private and public sectors. Although well-regarded and effective, in its most general form it is relatively complex and best implemented by specialists on major projects where time and expertise are both necessary and available.

What makes the Keeney and Raiffa model potentially demanding to apply is firstly that it takes uncertainty formally into account, building it directly into the decision support models and secondly that it allows attributes to interact with each other in other than a simple, additive fashion. It does not assume mutual independence of preferences. In certain circumstances, it can be important to build into the analysis one or both of these factors, but often in practice it may be better to ignore them in order to allow a simpler and more transparent decision support to be implemented more quickly, by a wider range of users and for a larger set of problem types.

4.7 Linear additive models

If it can either be proved, or reasonably assumed, that the criteria are preferentially independent of each other and if uncertainty is not formally built into the MCA model, then the simple linear additive evaluation model is applicable. The linear model shows how an option's values on the many criteria can be combined into one overall value. This is done by multiplying the value score on each criterion by the weight of that criterion, and then adding all those weighted scores together. However, this simple arithmetic is only appropriate if the criteria are mutually preference independent, a condition discussed in section 5.4.4.4. Most MCA approaches use this additive model, and it is the basis of the MCDA model developed in Chapter 6.

Models of this type have a well-established record of providing robust and effective support to decision-makers working on a range of problems and in various circumstances. They will form the foundation for the more detailed

²² Keeney, R.L. and Raiffa, H. (1976) *Decisions with Multiple Objectives: Performances and Value Trade-Offs*, Wiley, New York.

recommendations we shall give later. Some further details of the linear additive model are provided in Appendix 4.

However, as was argued earlier, the variety of circumstances in which decision support has been sought has led to the development of a range of different decision support models. A number of these will now be described and related to the basic MCDA model.

4.8 The Analytical Hierarchy Process

We will describe in some detail in Chapters 5 and 6 appropriate methods to assess the scores that are the basis of the performance matrix and for judging the weights in the linear additive model.

The Analytic Hierarchy Process (AHP) also develops a linear additive model, but, in its standard format, uses procedures for deriving the weights and the scores achieved by alternatives which are based, respectively, on pairwise comparisons between criteria and between options. Thus, for example, in assessing weights, the decision maker is asked a series of questions, each of which asks how important one particular criterion is relative to another for the decision being addressed.

The strengths and weaknesses of the AHP have been the subject of substantial debate among specialists in MCA. It is clear that users generally find the pairwise comparison form of data input straightforward and convenient. This feature is exploited in MCDA by the MACBETH approach to scoring and weighting (discussed in section 5.6) and the REMBRANDT approach (section 5.3). On the other hand, serious doubts have been raised about the theoretical foundations of the AHP and about some of its properties. In particular, the *rank reversal* phenomenon has caused concern. This is the possibility that, simply by adding another option to the list of options being evaluated, the ranking of two other options, not related in any way to the new one, can be reversed. This is seen by many as inconsistent with rational evaluation of options and thus questions the underlying theoretical basis of the AHP.

Details of the pairwise comparison process and some further assessment of the AHP model are set out in Appendix 5.

4.9 Outranking methods

A rather different approach from any of those discussed so far has been developed in France and has achieved a fair degree of application in some continental European countries. It depends upon the concept of outranking. The methods that have evolved all use outranking to seek to eliminate alternatives that are, in a particular sense, 'dominated'. However, unlike the straightforward dominance idea outlined in section 4.5, dominance within

the outranking frame of reference uses weights to give more influence to some criteria than others.

One option is said to outrank another if it outperforms the other on enough criteria of sufficient importance (as reflected by the sum of the criteria weights) **and** is not outperformed by the other option in the sense of recording a significantly inferior performance on any one criterion. All options are then assessed in terms of the extent to which they exhibit sufficient outranking with respect to the full set of options being considered as measured against a pair of threshold parameters. An explanation of precisely how outranking can identify a preferred option, or a set of preferred options for further investigation, is given in Appendix 6.

An interesting feature of outranking methods is that it possible, under certain conditions, for two options to be classified as 'incomparable' ('difficult to compare' is probably a better way to express the idea). Incomparability of two options is not the same as indifference between two options and might, for example, be associated with missing information at the time the assessment is made. This is not an unlikely occurrence in many decision making exercises. Building this possibility into the mathematical structure of outranking allows formal analysis of the problem to continue while neither imposing a judgement of indifference which cannot be supported nor dropping the option entirely, simply because information is not to hand.

The main concern voiced about the outranking approach is that it is dependent on some rather arbitrary definitions of what precisely constitutes outranking and how the threshold parameters are set and later manipulated by the decision maker.

The outranking concept does, however, indirectly capture some of the political realities of decision making. In particular it downgrades options that perform badly on any one criterion (which might in turn activate strong lobbying from concerned parties and difficulty in implementing the option in question). It can also be an effective tool for exploring how preferences between options come to be formed. However, on balance, its potential for widespread public use seems limited, notably in terms of many of the criteria set out in section 4.2.

4.10 Procedures that use qualitative data inputs

The view taken in this manual is that reliable and transparent support for decision making is usually best achieved using numerical weights and scores on a cardinal scale. There are some exceptions, for example application of dominance (Chapter 5) and use of models that approximate the linear additive model but are based on ranking of weights (Appendix 4). However, in general, it is a fair generalisation that the less precise the data inputs to any decision support procedure, the less precise and reliable will be the outputs that it generates.

Nonetheless, it is the case that decision makers working in government are frequently faced with circumstances where the information in the performance matrix, or about preference weights, consists of qualitative judgements. A number of methods exist to respond to this.

One group (discussed in Appendix 4) revolves around approximation to the linear additive model. In this respect they are relatively transparent, although they may involve significant amounts of data processing and, consistent with the fact that imprecise inputs rarely generate precise outputs to appraisal processes, usually require some extra assumptions to be made if, say, a single preferred option is to be identified, or even a ranking of options.

An alternative approach, largely developed in the Netherlands, has instead sought to develop procedures which amend outranking models in order to allow them to process imprecise, qualitative data inputs. Some examples are given in Appendix 7. They share many of the characteristics of (cardinal scale) outranking methods and have achieved only a limited degree of application, most often in urban and regional planning.

4.11 MCA methods based on fuzzy sets

A different response to the imprecision that surrounds much of the data on which public decision making is based has been to look to the newly developing field of fuzzy sets to provide a basis for decision making models. However, methods of this type are not yet widely applied.

Fuzzy sets attempt to capture the idea that our natural language in discussing issues is not precise. Options are 'fairly attractive' from a particular point of view or 'rather expensive', not simply 'attractive' or 'expensive'. Fuzzy arithmetic then tries to capture these qualified assessments using the idea of a membership function, through which an option would belong to the set of, say, 'attractive' options with a given degree of membership, lying between 0 and 1.

Building on assessments expressed in this way, fuzzy MCA models develop procedures for aggregating fuzzy performance levels using weights that are sometimes also represented as fuzzy quantities. Some examples are given in Appendix 8. However, these methods tend to be difficult for non-specialists to understand, do not have clear theoretical foundations from the perspective of modelling decision makers' preferences and have not yet established that they have any critical advantages that are not available in other, more conventional models. They are unlikely to be of much practical use in government for the foreseeable future.

4.12 Other MCA methods

The preceding sections have outlined some of the main types of MCA model that have been proposed as potentially applicable to public sector decision making. There are many others, some of which have a record of application, but many others which have not advanced significantly beyond the conceptual phase. Categories that have not been explicitly discussed but which are referred to in the MCA literature include methods based on Rough Sets, or on Ideal Points and several methods that are heavily dependent on interactive development, using specially constructed computer packages. For a variety of reasons, none of these is likely to find widespread application to mainstream public sector decision making.

A number of texts provide useful surveys of MCA methods, both those discussed in preceding sections and some of the more specialised procedures. Examples include Olson²³ and Yoon and Hwang.²⁴ The *Journal of Multi-criteria Decision Analysis* (John Wiley & Sons, Ltd) is a key source on new developments.

4.13 Application of MCA in government

There are many MCA methods, several of which have useful features that could justify their application. However, the role of this manual is to set out an approach that is broadly applicable across the range of government decisions and that fulfils the criteria set out in section 4.2 to the maximum possible extent. From these standpoints, an approach based firstly on establishing the performance matrix in a proper way, leading either to some limited analysis and choice by direct inspection (Chapter 5) and/or to the development of a linear additive (MCDA) model (Chapter 6) is the one recommended.

An important consideration in any decision making is risk and uncertainty. There are many ways in which risk can be handled in MCA. If it is a major feature in the decision concerned (for example in some nuclear safety applications) it can be built explicitly into multi-attribute utility models of the type developed by Keeney and Raiffa (Appendix 3).

For decision problems in general, it is more practicable not to try to model the uncertainty explicitly, but to undertake sensitivity testing of the rankings of options to changes in critical performance assessment inputs and/or criteria weights. Another possibility, facilitated by the use of the linear additive model, is to exploit the availability of risk analysis package 'add-ins' to standard spreadsheet programmes, to create profiles of possible overall performance level outputs reflecting estimates of the uncertainties surrounding key inputs.

²³ Olson, D. (1995) *Decision Aids for Selection Problems*, Springer Verlag, New York.

²⁴ Yoon, K.P., and Hwang, C-L (1995) *Multi-Attribute Decision Making*, Sage, Beverley Hills.

Chapter 5 Decisions without weights: construction of a performance matrix

5.1 The stages of a multi-criteria analysis

Chapter 4 introduced the concept of a performance matrix in which:

- each row describes one of the options that are being considered
- each column corresponds to a criterion, or 'performance dimension', which is considered important to the comparison of the different options, and
- the entries in the body of the matrix assess how well each option performs with respect to each of the criteria.

In this chapter, we outline how this performance matrix is created and how it fits into the overall scope of a multi-criteria analysis.

A full application of multi-criteria analysis normally involves eight steps.²⁵ These are set out in Figure 5.1.

This chapter describes steps 1 to 4 and step 7 for applications where there is no explicit numerical trade-off between criteria – in other words where steps 5 and 6 are omitted. These later steps are described in Chapter 6.

Multi-criteria analysis is described here as a cut and dried, step-by-step process. However, unless the user has applied the method to very similar problems in the past, it is more appropriate to envisage it as a guided exploration of a problem. Some of the steps will require detailed thought about issues surrounding the decision. It can be necessary to double back, re-visit earlier steps and revise them.

Much of the value derives from the thought that goes into the early steps.

²⁵ The Treasury's Green Book characterises public sector decision-making by the acronym ROAMEF (rationale, objectives, appraisal, monitoring, evaluation, feedback). The eight step approach to application of MCA in Figure 5.1 might be applied to appraisal, or to ex post evaluation, and can be used as part of a framework to monitor or to provide feedback on previous decisions.

Figure 5.1 Steps in a multi-criteria analysis

1. Establish the decision context. What are the aims of the MCA, and who are the decision makers and other key players?
2. Identify the options.
3. Identify the objectives and criteria that reflect the value associated with the consequences of each option.
4. Describe the expected performance of each option against the criteria. (If the analysis is to include steps 5 and 6, also 'score' the options, i.e. assess the value associated with the consequences of each option.)
5. 'Weighting'. Assign weights for each of the criteria to reflect their relative importance to the decision.
6. Combine the weights and scores for each of the options to derive an overall value.
7. Examine the results.
8. Conduct a sensitivity analysis of the results to changes in scores or weights.

We shall assume that the MCA is being implemented by a small team of public sector staff with a sound general knowledge of the area in which they are working and that their role is to provide advice on appropriate courses of action to more senior staff or to politicians. Where this is not the case, it may be necessary to make some small changes to the basic procedures.

It is worth noting that the application of MCA is not restricted to situations where the aim is to find only the single most appropriate option to follow through. Using just the steps of the MCA process covered in this chapter can be especially helpful when the requirement is to short-list a set of options for subsequent, more detailed investigation, or if the aim is to group options, in a transparent and defensible way, into categories ('urgent', 'less urgent', 'low priority', for example). Section 6.3.10 sets out a range of ways in which applications of the full MCDA procedure can serve the decision making process.

5.2 Step 1: Establishing the decision context

A first step is always to establish a shared understanding of the decision context.

The decision context is the whole panoply of administrative, political and social structures that surround the decision being made. Central to it are the objectives of the decision making body, the administrative and historical context, the set of people who may be affected by the decision and an identification of those responsible for the decision.

It is crucial to have a clear understanding of objectives. To what overall ambition is this decision seeking to contribute? MCA is all about multiple conflicting objectives. There are ultimately trade-offs to be made. Nonetheless, in applying MCA it is important to identify a single high level objective, for which there will usually be sub-objectives.

This may be thought of in terms of a value tree hierarchy, which is explored more fully in section 6.3.6 below.

To establish objectives (and criteria) we need to establish both who the decision-makers are (in order to establish objectives) and also people who may be affected by the decision.

A common component of this step can be to refer to underlying policy statements.

5.3 Step 2: Identifying options

Having established the decision context, the next step is to list the set of options to be considered.

It is unlikely, even with a new and unexpected problem, that the decision making group will arrive at the stage of formal MCA structuring without some intuition about options. Often in practice there will be ideas 'on the books', sometimes going back many years. Sometimes the problem will be an embarrassment of possibilities and it will be the role of the MCA in the first instance to provide a structured sifting of alternatives to identify a short-list, using basic data and quick procedures.

It is sometimes worth carrying out some informal sifting against established legal and similar restrictions. It is not worth considering and putting effort into gathering data about clearly infeasible propositions.

The first visit to step 2 may well not be the last, particularly in problems where there is a paucity of acceptable alternatives. The later steps of the MCA may demonstrate that none of the alternatives is acceptable and can serve to crystallise thoughts as to where the inadequacies lie. At this stage, fresh ideas and creative thought are needed. This will be informed by the MCA. For example, it may encourage a search for new options that combine the strong points of one existing option in some areas with the strong points of another in a different area.

The failure to be explicit about objectives, to evaluate options without considering what is to be achieved, led Keeney²⁶ to propose that starting with options is putting the cart before the horse. Options are important only for the value they create by achieving objectives. It might be better to consider objectives first, particularly when the options are not given and have to be developed.

5.4 Step 3: Identifying criteria and sub-criteria

5.4.1 OVERALL APPROACH

The criteria and sub-criteria are the measures of performance by which the options will be judged. A large proportion of the 'value-added' by a formal

²⁶ Keeney, R. L. (1992). *Value-Focused Thinking: A Path to Creative Decision making*, Harvard University Press, Cambridge, MA.

MCA process derives from establishing a soundly based set of criteria against which to judge the options.

Because the criteria serve as the performance measures for the MCA, they need to be operational. A measurement or a judgement needs to specify how well each option meets the objectives expressed by the criteria. We shall return to this later, but a question to be borne in mind in developing the set of criteria is **“Is it possible in practice to measure or judge how well an option performs on these criteria?”**

5.4.2 PROCEDURES TO DERIVE CRITERIA

Whether in a decision making team or as an individual, an effective way to start the process of identifying criteria is first briefly to recapitulate step 1 and then to brainstorm responses to the question **“What would distinguish between a good choice and a bad one in this decision problem?”**

Responses should all be noted down uncritically, perhaps on whiteboards if in a group context.

Interest group perspective(s) may be important. One way to include them is directly to involve the affected parties in some or all stages of the MCA. This might be appropriate, for example, in some local planning issues. A second approach is to examine policy statements and secondary information sources from the various interest groups and to analyse these to derive criteria to reflect their concerns. A third, if suitable experience resides within the decision making team, is to encourage one or more of its members to role-play the position of key interest groups, to ensure that this perspective is not overlooked when criteria are being derived.

Often, both decision-maker objectives and interest group viewpoints may be articulated in broad-brush terms. For example, a criterion like environmental impact might be suggested. In many circumstances, assessing options against such a broad criterion may prove difficult, even though the notion of environmental impact may be important. Vague criteria are generally not useful in MCA, any more than they are in CBA.²⁷

Typically, in the process of eliciting criteria, after an initial hesitation, suggestions come thick and fast, until eventually the process slows and dries up. At the end of a relatively short period, it is normal to have a substantial list of potential criteria.

The number of criteria should be kept as low as is consistent with making a well-founded decision. There is no ‘rule’ to guide this judgement and it will certainly vary from application to application. Large, financially or otherwise important choices with complex technical features (such as a decision on where to locate a nuclear waste facility) may well have upwards of a hundred criteria. More typical, however, is a range from six to twenty.

²⁷ One approach is to think in terms of a value tree, or hierarchy of criteria. See section 6.3.6 below.

5.4.3 GROUPING CRITERIA

It can be helpful to group together criteria into a series of sets that relate to separate and distinguishable components of the overall objective for the decision. This is particularly helpful if the emerging decision structure contains a relatively large number of criteria (say eight or more).

The main reasons for grouping criteria are: (a) to help the process of checking whether the set of criteria selected is appropriate to the problem (see section 5.4.4); (b) to ease the process of calculating criteria weights in large MCDA applications (see section 6.2.10), when it can sometimes be helpful to assess weights firstly within groups of related criteria, and then between groups of criteria; and (c) to facilitate the emergence of higher level views of the issues, particularly how the options realise trade-offs between key objectives.

For both these reasons, grouping criteria ('clustering' them or 'structuring the value tree' in the language of a full MCDA, as in section 6.2.6) is an important part of an MCA. However, there are few formal guidelines to determine what is a 'good' structure and what is 'bad'. Most experienced decision analysts see problem structuring as a skill that is acquired primarily through practical experience. For most large problems, there is arguably no unambiguously correct structure or grouping of criteria.

An acceptable structure is simply one that reflects a clear, logical and shared point of view about how the many criteria that may be relevant to an MCA assessment can be brought together into coherent groups, each of which addresses a single component of the overall problem. For example, in assessing forms of medical intervention for a given condition, one group of criteria may relate to the patient's experience (speed with which treatment could be obtained, length of stay in hospital, degree of discomfort, effectiveness of treatment, etc.). Often the criteria in an MCA reflect individual measurable indicators of performance relative to the issue at hand, whereas the groups of criteria reflect sub-objectives to the single main objective that underlies the MCA process.

While knowledge of the domain of the particular problem will often give very clear guidance as to what are clear and helpful groups of criteria, there can be room for debate. For example, should criteria relating to the time to the end of the treatment (speed of admission, length of stay) constitute one sub-objective, with criteria reflecting the experience of the treatment itself being placed in their own cluster? To some extent, such debate is helpful and to be expected. It is one way in which the decision makers explore the problem to be solved and come to a shared understanding of its characteristics and what factors should drive their choice.

Especially in some complex and contentious public sector decisions, it is likely that different stakeholder groups, because of their very different ways of framing the problem, may have substantial difficulty in sharing the same grouping of criteria. It may need prolonged negotiation and a good deal of tact and ingenuity to arrive at a shared structure. For a useful discussion of

structuring in the context of transport of radioactive waste in the UK, see Brownlow and Watson (1987).²⁸

5.4.4 ASSESSMENT OF THE PROVISIONAL SET OF CRITERIA

Before finalising the choice of criteria the provisional set needs to be assessed against a range of qualities.

5.4.4.1 *Completeness*

Have all important criteria been included? This needs some care as it is not necessarily obvious from the beginning what the important criteria are.

For this process, the value tree, if one has been sketched out, can be a valuable aid (see section 6.3.6). First, the team may review the set of major sub-headings from the criteria groups and ask **Have we overlooked any major category of performance?** Secondly, **within** each head, it can ask, **With regard to this area of concern, have we included all the criteria necessary to compare the options' performance?** Thirdly, **Do the criteria capture all the key aspects of the objectives that are the point of the MCA?**

5.4.4.2 *Redundancy*

Are there criteria which are unnecessary? In principle, criteria that have been judged relatively unimportant or to be duplicates should have been removed at a very early stage, but it is good practice to check again.

The MCA team may also wish to delete a criterion if it seems that all the available options are likely to achieve the same level of performance when assessed against it. If this were the case, then omitting it would not affect any ranking of options and would economise on analytical input. However, omission on these grounds should be approached with care. First, it has not yet been formally assessed how well each option will show up on the criterion concerned. Secondly, it may be that at a later stage new options come into consideration which do not exhibit this behaviour, especially in MCA systems that may be used by delegated groups and/or to address a series of different problems.

5.4.4.3 *Operationality*

It is important that each option can be judged against each criterion. The assessment may be objective, with respect to some commonly shared and understood scale of measurement, like weight or distance. Optionally, it can be **judgmental**, reflecting the subjective assessment of an expert. A strength of MCA is its ability to accommodate and use simultaneously both forms of assessment of options. In either case, however, the criterion must be defined clearly enough to be assessed.

²⁸ S.A. Brownlow and S.R. Watson, "Structuring multi-attribute value hierarchies", *Journal of the Operational Research Society*, vol. 38, 1987, pp.309–317.

It can sometimes be helpful to break a criterion down into a further sub-level of more explicitly defined criteria, if assessment at a particular level is problematic.

5.4.4.4 Mutual independence of preferences

Straightforward applications of MCA require that preferences associated with the consequences of the options are independent of each other from one criterion to the next. The key idea is simple: can you assign preference scores for the options on one criterion without knowing what the options' preference scores are on any other criteria? If the answer is yes, then this criterion is preference independent of the others. The question then is asked for each of the remaining criteria in turn. If the answer is always yes, then the criteria are considered to be mutually preference independent. This condition has to be met if the sum of weighted averages is to be used to combine preference scores across criteria, and this is true for all MCA approaches, whether they recognise it formally or not.

Preferences are not always mutually independent. For example, the enjoyment a person gets from consuming a trifle may not be the sum of the amount of jelly, custard, sponge, etc. it contains, but be related in some way to the proportions in which they are combined. If this is the case, a simple weighted sum of the amounts of jelly, custard and so forth contained in a set of option trifles will not in general reproduce the preference ranking that the individual has for the trifles.

In practical terms, the preferential independence question may be approached by asking, for each criterion, whether the preference scores of an option on one criterion can be assigned independently of knowledge of the preference scores on all the other criteria.

If the answer is no, then it may be necessary to use more complex models for combining scores across criteria. However, two simpler approaches may be possible. The first is to combine the two criteria that are not preference independent of each other into a single criterion, which captures the common dimension of value. This will be effective provided that the new criterion is itself preference independent of the remaining criteria. The second approach is to recognise that options often have to satisfy a minimum acceptable level of performance for them to be considered; options falling below any minimum level are rejected outright because better performance on other criteria can't compensate. This hurdle usually guarantees preference independence of the criteria; all options fall at or above the minimum level of performance, so that preference on any given criterion is unaffected by preference on the others. If preference independence is still violated, then more advanced MCA procedures must be adopted.

5.4.4.5 Double counting

Public sector decisions can be particularly prone to double counting, especially of effectiveness or benefits. This stems, for example, from a desire to set out the distribution of effects on different parts of the population. As a

consequence, it is quite easy for the same basic impact to be recorded more than once in a performance matrix.

As with CBA, double counting should not be allowed in MCA, since double-counted effects are likely to be given more weight in the final overall decision than they deserve. The checks set out in sections 5.4.2 and 5.4.3.2 are designed to stop double counting from occurring.

On occasions, however, it is desirable to present the same basic effect from more than one point of view, so that the overall context of the decision is fully understood by those involved. For example, both numbers of accidents saved and the money value of accident cost savings are sometimes recorded as separate items in appraisal of transport schemes. It is important, however, when moving from this multi-perspective form of presentation of options on to the process of choice between options, that the potential for any double counting is recognised and the final performance matrix used for decision making is suitably amended to remove it.

Sometimes, what appears to be double counting isn't. For example, in prioritising newly discovered compounds for possible development into effective drugs, a pharmaceutical company will be guided in prioritising its investments by the extent to which the compounds will meet unmet medical need. If a disease state is widespread and serious, then the unmet medical need will be great, and the commercial value is potentially large. Thus, two separate criteria, commercial value and medical need, would appear to be double counting. That may be true for companies only concerned with generic drugs, but for the large pharmaceutical companies this might not be the case.

A good test is to ask the following question: "Two compounds, A and B, will cost the same to develop, are expected to yield the same financial return, and are identical on all other criteria except that A meets a greater unmet medical need than B. Do you prefer A or B, or are you indifferent?" If the answer is a preference for A, then there must be more to the value associated with A than its expected commercial value. Exploring the reasons for the preference will uncover additional criteria. For example, the pharmaceutical company Bristol-Myers Squibb includes 'extending and enhancing human life' as part of its corporate mission, and so compound A might be seen as fulfilling that mission better than B. Realising the mission provides value to the company in addition to financial value. Further, A might be more innovative, and so contribute to the image of the company as a world leader in cutting-edge science, thus attracting high-quality scientists to work for the organisation. In short, many non-financial values might be invoked, suggesting that medical need can be partitioned into separate aspects of value, commercial or financial value, and one or more criteria representing non-financial value. Because innovative products that fulfil the corporate mission and create a favourable image for the company are also likely to be commercially successful, the scores of options on financial and non-financial criteria will be correlated, yet the criteria represent separate aspects of value. Preferences are independent, and there is no double counting.

This example shows that a judgement about double counting cannot be made on an objective basis. It is necessary to understand the values that the organisation brings to the appraisal. Fortunately, checking for independence of preferences will reveal double counting: if criteria X and Y really reflect the same value, then when scoring options on Y, one will wish to look at scores given to X.

5.4.4.6 Size

An excessive number of criteria leads to extra analytical effort in assessing input data and can make communication of the analysis more difficult. A final check to ensure that the structure is no larger than it needs to be is useful at this stage.

In a full application of MCA, criteria are explicitly weighted. However, in the absence of weighting, if there is any possibility of informal judgements being made by scanning the performance matrix, it is wise at this stage to ensure that marked inconsistencies between the number of criteria and the likely importance of the topics they reflect are, if possible, eliminated. If this is not practicable, then particular care needs to be taken to prevent the imbalance distorting people's interpretation of the matrix.

5.4.4.7 Impacts occurring over time

Many public sector decisions concern expenditures to be undertaken now that will have impacts occurring over several subsequent years. Although this can present some difficulties in aggregating all effects into a single measure, with monetary-based techniques discounting is a reasonably well established procedure for aggregation. In MCA there is no single equivalent technique, although in principle the conventional discounting of money values can be accommodated and it can also be applied to physical impact indices other than monetary value. The reasons why these are not generally done are probably more cultural than substantive. Certainly, good decision facilitating practice would ensure that participants in any decision making exercise had their attention drawn to time-differentiated impacts and gave thought to how these were to be consistently accommodated in the assessment.

If a target completion date is an important consideration, it can be modelled as a separate criterion, with a non-linear value function. Options that are expected to deliver on time get good scores, those expected to deliver slightly late receive lower scores, and very late ones get zeros. Time has to be included in the definition of many other criteria so that temporary consequences can be distinguished from permanent ones. This is usually done by being explicit about the time horizon over which the consequences are being valued. Time horizons may differ from one criterion to the next, eg separately identifying short-term and long-term health effects.

A further possibility would be to use some other principle of giving less importance to impacts in the long-run future. Alternatively there are approaches supported by some environmentalists for giving greater influence to longer term impacts. Finally, it would be possible to carry out an MCA

within an MCA, using expert judgements to assess the weights to be applied to impacts occurring in different future time periods.

The number of occasions when discounting or other analytical ways of tackling time-distributed impacts is needed in MCA applications is likely to be relatively limited. There is relatively little published guidance on time preference issues in MCA, although the chapter by John Meyer in Keeney and Raiffa gives a good introduction. The key point for users of MCA to bear in mind is to ensure that all assessments of criteria are made on the same basis. Thus if some impacts are one-off whereas others are repeated (perhaps with different types of time profile) then these differences need to be explicitly recognised in the scores which are given to alternatives on the relevant criteria.

5.5 Step 4 and beyond (without scoring and weighting)

5.5.1 THE PERFORMANCE MATRIX WITHOUT SCORING AND WEIGHTING

A basic MCA will present the decision maker with the performance matrix itself. The task for the decision maker is then to study the matrix, and come to a view on the ranking of the options – probably assisted by some supplementary advice from those who constructed the matrix on their views of how the information should be interpreted.

Matrices are commonly presented in this way by consumer journals and magazines, in comparing for example electrical goods, sports equipment, computer software, or other consumer goods. For this purpose they are especially suitable because they are addressing typically not one but hundreds or thousands of decision makers, each with his or her own priorities.

The measures used in performance matrices of this kind are often qualitative descriptions (for example of styling), or natural units (such as price or length), or sometimes a crude numerical scale (eg number of stars), or even a scale of 0 to 100.

For government applications the use of 0 to 100 numerical scales is not recommended if the analysis is not to proceed to the numerical analysis of stages 5 and 6. The extra work entailed in producing such scales can all too easily be counterproductive, by giving the intuitive but incorrect message that the scores can then be added together.

Even if the matrix is confined to qualitative description, natural units and very simple scales (such as stars) it is advisable to try to use similar numbers of criteria within each major sector of the value tree (the balance issue is discussed in Section 5.4.3.6 above).

It is also worth considering the use of supplementary presentations of the data, such as graphs, to help encourage people to think about the data in different ways and avoid giving undue weight to some factors relative to others.

5.5.2 JUDGEMENTS BETWEEN OPTIONS WITHOUT SCORING AND WEIGHTING

To what extent does a performance matrix alone allow any comparison of options? The basic answer is only a little. What is perhaps just as important is to be clear about what types of comparison may **not** be made, and why.

5.5.2.1 *Dominance*

First, it is possible to examine the set of options to check for the presence of dominance. As noted in Section 4.5, one option dominates another if it performs at least as well on all criteria and is strictly better on at least one criterion.

Assuming that all the estimates of criteria scores are accurate, if option A dominates option B, then B cannot be the single best one available. Thus, if the purpose of the MCA is to recommend a single best option, B may be removed from consideration. If the purpose is short-listing, then it is possible, but rather unlikely, that a dominated option would be taken through to a later stage in the selection process. Logically it would only make sense to do so if it was thought that new information about options might come to light, that some of the criteria scores might be inaccurate, or if there was some possibility that the dominating option (A) might cease to be available.

For screening, being dominated might exclude an option from further consideration, depending on the number of options required for later consideration and the strength of the others available, but dominance says only that B must rank lower than A.

Finally, it can be noted that dominance is transitive. If A dominates B, and B dominates C, then A will always dominate C and this dominance does not need to be directly checked.

In practice, dominance is rare. The extent to which it can help to discriminate between options and so to support real decisions is correspondingly limited.

5.5.2.2 *Other approaches*

Dominance is limited in the extent to which it can differentiate between options specifically because it makes no assumption at all about the relative importance of criteria (the different columns), nor does it employ any supplementary information beyond that directly displayed in the performance matrix. It is a common (although often sub-conscious) intuitive mis-use of the performance matrix to either:

- (a) add recorded performance levels across the rows (options) to make some holistic judgement between options about which ones are better
- (b) eliminate (or prioritise) options that record weak (or strong) performance levels on particular criteria.

In the first case, the implication is that all criteria contribute with equal importance to options' overall performance, when this has not been established. In (b) the same error is potentially made (in deciding the criteria where strong or weak performances are identified) and there may additionally be a degree of arbitrariness in deciding the thresholds for recognising a special level of performance.

In addition to dominance, there are a limited number of *non-compensatory* MCA procedures that may be applied, using supplementary information of various sorts beyond that in the performance matrix. These are discussed in Appendix 2. However, they are of limited value in practice.

5.5.2.3 *The limitations of human judgements*

Research on human judgements and decision making²⁹ shows that the simplifications which we make to enable us to deal with complex problems sometimes do not work well. We are inclined for example to be biased in our assessments of alternatives that can more readily be linked to what is familiar (the 'representativeness heuristic'), and to be unduly influenced by recent, memorable, or successful experience (the 'availability heuristic').

MCA techniques are designed to help overcome the limitations by imposing a disciplined structure which directs attention to criteria in proportion to the weight which they deserve.

The development of a performance matrix is an important step in this direction, but it is limited because a subjective interpretation of the matrix is still prone to many of these well documented distortions of human judgement, as well as the intuitive processing errors set out in Section 5.5.2.2.

In practice, the extent to which options can be compared using *non-compensatory* methods is strictly limited. The alternatives at this stage are either to end the MCA, reverting to an informal treatment of the decision, for which the performance matrix simply provides basic factual information in a more considered and coherent way than might otherwise have been the case, or to move on to a formal, *compensatory* MCA.

The next section discusses how a performance matrix can be scaled in numerical terms as a precursor to the more sophisticated *compensatory* MCA described in Chapter 6.

5.6 Step 4: Assessing performance levels (with scoring)

The first consideration in setting up consistent numerical scales for the assessment of criteria is to ensure that the sense of direction is the same in all cases, so that (usually) better levels of performance lead to higher value scores. This may mean a reversal of the natural units. For example, access

²⁹ For example, Bazerman, M H (1998) *Judgement in Managerial Decision Making* John Wiley, New York, Fourth Edition.

to a facility might be recorded in terms of distance to the nearest public transport, where the natural scale of measurement (distance) associates a low number with a good performance.

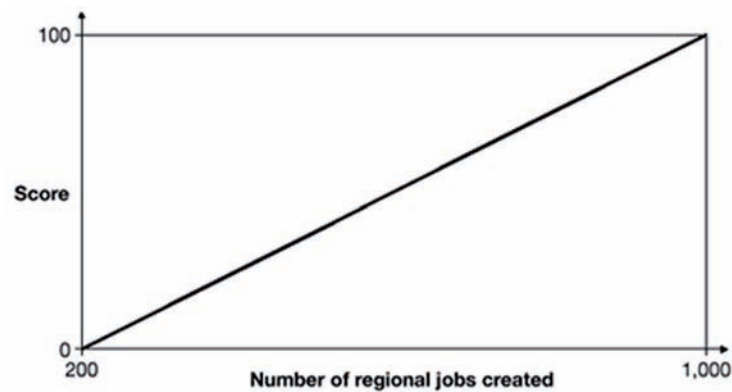
It is conventional to allot a value score to each criterion between 0 and 100 on an interval scale. The advantage of an interval scale is that differences in scores have consistency within each criterion, although it does not allow a conclusion that a score of 80 represents a performance which on any absolute standard is five times as good as a score of 16 (which would require a ratio scale of measurement). The 'ruler' which the scoring scale represents is good only within the confines of this particular MCA. However, when combined with appropriately derived importance weights for the criteria, the use of an interval scale measurement does permit a full MCA to be pursued.

The first step in establishing an interval scale for a criterion is to define the levels of performance corresponding to any two reference points on the scale, and usually the extreme scores of 0 and 100 would be used. One possibility (global scaling) is to assign a score of 0 to represent the worst level of performance that is likely to be encountered in a decision problem of the general type currently being addressed, and 100 to represent the best level. Another option (local scaling) associates 0 with the performance level of the option in the currently considered set of options which performs least well and 100 with that which performs best.

The choice between local and global should make no difference to the ranking of options. An advantage of global scaling is that it more easily accommodates new options at a later stage if these record performances that lie outside those of the original set. However it has the disadvantages of requiring extra, not necessarily helpful judgements in defining the extremes of the scale and, as will be seen in the next chapter, it lends itself less easily than local scaling to the construction of relative weights for the different criteria.

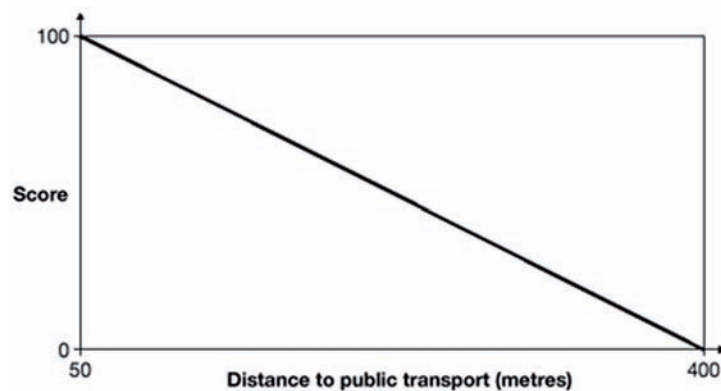
Once the end points are established for each criterion, there are three ways in which scores may be established for the options.

The first of these uses the idea of a value function to translate a measure of achievement on the criterion concerned into a value score on the 0 – 100 scale. For example, if one criterion corresponds to number of regional full-time jobs created and the minimum likely level is judged to be 200 and the maximum 1,000, then a simple graph allows conversion from the natural scale of measurement to the 0 – 100 range required for the MCA. This is shown in Figure 5.2.

Figure 5.2 Regional jobs

For any option, its score on the regional job creation criterion is assessed simply by reading off the vertical axis the score corresponding to the number of jobs created, as measured on the horizontal axis. Thus an option that creates 600 jobs, say, scores 50.

Where higher measurements on the scale of natural units correspond to worse rather than better performance, the slope of the function mapping achievement level on to 0–100 score is simply reversed, as in Figure 5.3.

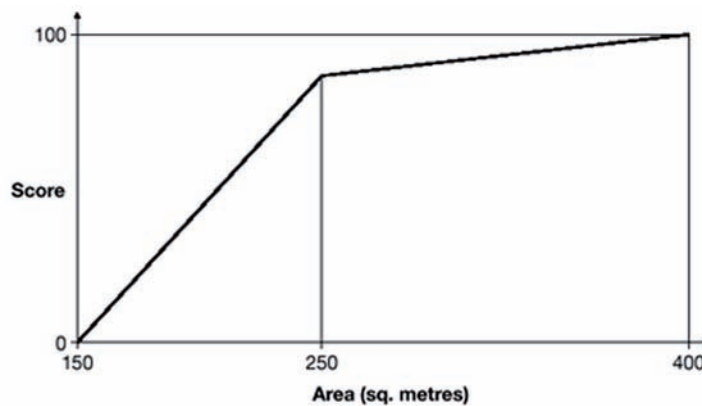
Figure 5.3 Distance to public transport

The value functions used in many MCA applications can for practical purposes be assumed to be linear. However, on some occasions it may be desirable to use a non-linear function. For example, it is well known that human reaction to changes in noise levels measured on a decibel scale is non-linear. Alternatively, there are sometimes thresholds of achievement above which further increments are not greatly appreciated. For example, in valuing office area, it may well be that increments above the absolute minimum initially lead to substantially increased judgements of value on room size, but after an acceptable amount of space is available, further marginal increments are valued much less highly. This is illustrated in Figure 5.4. In this case, the judgement is that, once the area reaches about 250 square metres, further increments add less value.

The second approach to scoring performance on an interval scale is direct rating. This is used when a commonly agreed scale of measurement for the criterion in question does not exist, or where there is not the time nor the resources to undertake the measurement. Direct rating uses the judgement of an expert simply to associate a number in the 0–100 range with the value of each option on that criterion.

Because these scores are being assessed on an interval scale of measurement, relationships between the differences in options' scores do have meaning and it is important to check that the judgements being made are consistent in this respect. Specifically, a difference of (say) 20 points should reflect an improvement in assessed value which is exactly half that measured by a difference of 40 points.

Figure 5.4 Diminishing returns to area



The use of direct rating judgements in MCA can pose some problems of consistency in circumstances where the procedure is to be applied by different people, eg, where some decision making responsibility is delegated to regional offices. The simplest way to encourage consistency is to provide a set of examples or scenarios with suggested scores associated with them.

Another issue to bear in mind with direct rating is that sometimes those with the most appropriate expertise to make the judgements may also have a stake in the outcome of the decision. Where this is the case, there is always the danger that their rating judgements may (perhaps unconsciously) be influenced by factors other than simply the performance of the options on the criterion being assessed. Ideally, such judgements should come from individuals who are both expert and disinterested. If this cannot be managed, then it is important to be aware of the possibility of some bias creeping in and, for example, to apply sensitivity testing to the scores at a later stage as a means of checking the robustness of the outcome of the analysis.

A third approach to scoring the value of options³⁰ on a criterion is to approach the issue indirectly, by eliciting from the decision maker a series of verbal pairwise assessments expressing a judgement of the performance

³⁰ Each of the methods mentioned here can also be used to establish the relative weights to be given to criteria in the full MCDA procedure set out in Chapter 6.

of each option relative to each of the others. The Analytic Hierarchy Process (AHP) does this (Chapter 4 and Appendix 5). Alternatives are to apply REMBRANDT (Appendix 5) or MACBETH (see, eg, Bana e Costa and Vansnick, 1997, Bana e Costa et al, 1999).³¹

To illustrate, the MACBETH procedure asks decision makers to assess the attractiveness difference between each pair of options as one of:

C_1 very weak difference

C_2 weak difference

C_3 moderate difference

C_4 strong difference

C_5 very strong difference

C_6 extreme difference

Once all the required pairwise comparisons are made, a series of four computer programmes processes these data to calculate a set of scores for the options, on a 0–100 scale, which are mutually consistent with the full set of stated pairwise judgements. If, as can happen, there are inconsistencies within the judgements, such that a compatible set of scores cannot be computed from them, the programmes guide the decision maker through steps to amend the inputs until consistent scores are obtained.

Following one or other of these procedures through one by one for each of the criteria provides the full set of value scores on which any compensatory MCA must be based. Applications which apply formal numerical comparisons to options are covered in Chapter 6.

³¹ Bana e Costa, C.A. and Vansnick, J.C. Applications of the MACBETH approach in the framework of an additive aggregation model, *Journal of Multi-criteria Decision Analysis*, vol. 6, pp. 107–14, 1997, and Bana e Costa, C.A., Ensslin, L., Correa, E.C. and Vansnick, J.C. Decision support systems in action: integrated application in a multi-criteria decision aid process, *European Journal of Operational Research*, vol. 113, pp.315–35, 1999.

Chapter 6 Multi-criteria decision analysis

6.1 What is MCDA?

A form of MCA that has found many applications in both public and private sector organisations is multi-criteria decision analysis, or MCDA for short (also known as multi-attribute decision analysis, or MADA). This chapter explains what MCDA is and then outlines what is required to carry out such an analysis.

MCDA is both an approach and a set of techniques, with the goal of providing an overall ordering of options, from the most preferred to the least preferred option. The options may differ in the extent to which they achieve several objectives, and no one option will be obviously best in achieving all objectives. In addition, some conflict or trade-off is usually evident amongst the objectives; options that are more beneficial are also usually more costly, for example. Costs and benefits typically conflict, but so can short-term benefits compared to long-term ones, and risks may be greater for the otherwise more beneficial options.

MCDA is a way of looking at complex problems that are characterised by any mixture of monetary and non-monetary objectives, of breaking the problem into more manageable pieces to allow data and judgements to be brought to bear on the pieces, and then of reassembling the pieces to present a coherent overall picture to decision makers. The purpose is to serve as an aid to thinking and decision making, but not to take the decision. As a set of techniques, MCDA provides different ways of disaggregating a complex problem, of measuring the extent to which options achieve objectives, of weighting the objectives, and of reassembling the pieces. Fortunately, various computer programs that are easy to use have been developed to assist the technical aspects of MCDA, and these are set out in the Software review.

The first complete exposition of MCDA was given in 1976 by Keeney and Raiffa,³² whose book is still useful today. They built on decision theory, which for most people is associated with decision trees, modelling of uncertainty and the expected utility rule. By extending decision theory to accommodate multi-attributed consequences, Keeney and Raiffa provided a theoretically sound integration of the uncertainty associated with future consequences and the multiple objectives those consequences realise.

The main assumption embodied in decision theory is that decision makers wish to be coherent in taking decisions. That is, decision makers would not

³² Keeney, R. L., & Raiffa, H. (1976) *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*, John Wiley, New York, reprinted, Cambridge University Press, 1993.

deliberately set out to take decisions that contradict each other. No-one would place several bets on the outcome of a single race such that no matter which horse won they were certain to lose money. The theory expands on this notion of coherence, or consistency of preference, and proposes some simple principles of coherent preference, such as the principle of transitivity: if A is preferred to B, and B to C, then A should be preferred to C, which is a requirement if preference is to be expressed numerically. By treating these rather obvious principles as axioms it is possible to prove non-obvious theorems that are useful guides to decision making. A parallel can be found in the study of geometry. Simple principles like 'The shortest distance between two points is a straight line' are combined using the rules of logic to prove theorems that are not obvious, like the Pythagorean principle, that the square of the hypotenuse equals the sum of the squares of the other two sides.

The first two theorems establish a logical equivalence between coherent preference and number systems. If preferences are coherent, then two sorts of measures follow logically: probability and utility, both associated with the consequences of decisions. The first theorem establishes the existence of probabilities: numbers which capture the likelihood that consequences will occur. The second theorem shows the existence of utilities: numbers which express the subjective value of the consequence and the decision maker's risk attitude.

The third theorem provides a guide to taking decisions: choose the course of action associated with the greatest sum of probability-weighted utilities. That is the expected utility rule, which has existed in various guises for over 200 years. To apply the expected utility rule, assess a probability and utility for each possible consequence of a course of action, multiply those two numbers together for each consequence, and add those products to give the expected utility for that course of action. Repeat the process for each course of action, and choose the action associated with the largest expected utility. That description sounds rather dry and impractical, but decision theory gave birth to the applied discipline of decision analysis.³³ Thousands of decision analyses have been successfully carried out since the 1960s in all aspects of organisational life.

Keeney and Raiffa extended the set of axioms so that decisions with multiple objectives could be analysed. In practice, MCDA is applied to help decision makers develop coherent preferences. In other words, coherent preferences are not assumed to start with, but the approach helps individuals and groups to achieve reasonably coherent preferences within the frame of the problem at hand. Once coherent preferences are established, decisions can be taken with more confidence.

The years following the publication of Keeney and Raiffa's book saw increasing numbers of applications of MCDA in both private and public sectors. Many of these are referenced in the bibliography to this manual. The

³³ A term first coined by Professor Ronald Howard in 1966. The first complete exposition was given by Howard Raiffa in his 1968 book, *Decision Analysis: Introductory Lectures on Uncertainty*.

use of MCDA by various governmental agencies in the United States, at local, state and federal level, is now widespread. The approach has also withstood challenges of its results in courts of law and inquiries. The audit trail left by a well-conducted MCDA suits the climate of freedom of information in the United States, though it has not always been favoured by those who wished to take very different decisions from those recommended by the analysis.

A notable example was the analysis of alternative sites for the disposal of nuclear waste in the United States. Five potential sites were analysed using MCDA, which resulted in an overall ranking of the sites. The US Department of the Environment's subsequent announcement of three sites for further investigation included the sites ranked first, third and fifth by the MCDA rather than the first three.³⁴ Puzzled by this, Keeney³⁵ conducted a new analysis whose purpose was to find the best three combinations of sites for further investigation, for it would not be cost effective to investigate simultaneously any two very similar sites. Keeney's analysis of this more complex portfolio problem shows that a sequential characterisation strategy would be more cost efficient, but still excluded the originally fifth-rated site. Reaction to the DOE announcement was swift and drastic. Congress and the House of Representatives initiated an investigation into the DOE's decision process and 46 lawsuits were filed against the DOE charging them with violations of federal laws in their selection process. The investigation supported the MCDA analysis, but concluded that the DOE's decision process was flawed.

Several lessons were drawn from this experience in a subsequent paper³⁶ which looked at the whole sequence of events. For projects of major public concern, 'it is crucial to obtain inputs from a variety of professionals and to have the implementation of the methodology monitored and routinely reviewed by independent experts.' The authors recommended including objectives of key interest groups, making all value judgements explicit, analysing as many crucial problem complexities as possible, obtaining information from independent professionals, communicating all aspects of the analysis to interested parties and individuals, and conducting an independent review. While all these will not be appropriate for more modest projects, it is notable that the recommendations focus on social issues. MCDA is not simply a technical process. Its successful implementation depends crucially on effective design of social processes by which the analysis is structured and conducted.

³⁴ Merkhofer, M. W. and R. L. Keeney (1987) "A multiattribute utility analysis of alternative sites for the disposal of nuclear waste," *Risk Analysis* 7(2): 173–94.

³⁵ Keeney, R. L. (1987) "An analysis of the portfolio of sites to characterize for selecting a nuclear repository," *Risk Analysis* 7(2): 195–218.

³⁶ Keeney, R. L. and D. von Winterfeldt (1988) "The analysis and its role for selecting nuclear repository sites", *Operational Research '87*. G. K. Rand, Elsevier Science Publishers B.V. (North Holland).

6.2 Stages in MCDA

MCDA can be used either retrospectively to evaluate things to which resources have already been allocated, or prospectively to appraise things that are as yet only proposed. Thus, in the following explanations of MCDA, there is no need to distinguish these two uses, though in practice the approach will be realised differently.

In discussing the application of the eight-step process introduced in Chapter 5, and illustrated in Figure 5.1, we divide some of these steps further in the explanation of the application of MCDA in this chapter. The more detailed steps are shown in Figure 6.1.

The sections to follow describe what has to be done at each step, leaving details of how to execute the steps to the real examples in Chapter 7. Some of the process is technical, but equally important is organising the right people to assist at each stage, and some suggestions about these social aspects will be given in this Chapter. A simple example of MCDA will be used to illustrate the stages in selecting a toaster for Fred Jones's family. This modest decision problem would hardly require a full MCDA, but it does provide an illustration unencumbered by the detail and difficulties met in real applications. In the next chapter, real examples will be given. Fred's MCDA appears in the boxes, and the reader wanting a quick introduction to MCDA could just read the boxes.

Chapter 5 has described steps 1 to 4 and these are developed in more detail in this chapter, where the construction of weights (step 5) and their subsequent combination with scores is described for the first time. The use of weights presents two kinds of challenge. One is the need for exceptional care to ensure logical consistency between the ways in which weights and scores are constructed, as explained in section 6.2.10. The other challenge, in some cases, is the largely social problem of handling widely different value judgements of different contributors. This is illustrated in the NIREX case study in Chapter 7.

Figure 6.1 Applying MCDA: Detailed steps

- 1. Establish the decision context.**
 - 1.1 Establish aims of the MCDA, and identify decision makers and other key players.
 - 1.2 Design the socio-technical system for conducting the MCDA.
 - 1.3 Consider the context of the appraisal.
- 2. Identify the options to be appraised.**
- 3. Identify objectives and criteria.**
 - 3.1 Identify criteria for assessing the consequences of each option.
 - 3.2 Organise the criteria by clustering them under high-level and lower-level objectives in a hierarchy.
- 4. 'Scoring'. Assess the expected performance of each option against the criteria. Then assess the value associated with the consequences of each option for each criterion.**
 - 4.1 Describe the consequences of the options.
 - 4.2 Score the options on the criteria.
 - 4.3 Check the consistency of the scores on each criterion.
- 5. 'Weighting'. Assign weights for each of the criterion to reflect their relative importance to the decision.**
- 6. Combine the weights and scores for each option to derive an overall value.**
 - 6.1 Calculate overall weighted scores at each level in the hierarchy.
 - 6.2 Calculate overall weighted scores.
- 7. Examine the results.**
- 8. Sensitivity analysis.**
 - 8.1 Conduct a sensitivity analysis: do other preferences or weights affect the overall ordering of the options?
 - 8.2 Look at the advantage and disadvantages of selected options, and compare pairs of options.
 - 8.3 Create possible new options that might be better than those originally considered.
 - 8.4 Repeat the above steps until a 'requisite' model is obtained.

6.2.1 ESTABLISH AIMS OF THE MCDA, AND IDENTIFY DECISION MAKERS AND OTHER KEY PLAYERS

What is the purpose of the MCDA? Get this wrong and you can provide a wonderful analysis for the wrong problem. That's not to say the purpose stays fixed throughout the analysis. As an MCDA progresses new features are often revealed and new issues raised, which may signal a change or shift of aims. Still, the MCDA has to start somewhere, and a statement of initial aims is crucial to formulating the successive stages. After all, MCDA is about determining the extent to which options create value by achieving objectives, and at this stage you face two options: doing the MCDA or not. Choosing to carry out the MCDA means that someone judged the analysis to provide relatively more value than not doing it.

Clarity about the aims of the MCDA helps to define the tasks for subsequent stages and keeps the analysis on track.

Aims of the toaster MCDA:

Fred Jones has an old toaster he bought for less than £10 many years ago, and it now toasts unevenly. His youngest daughter Zoe burned her finger when she touched the side of the toaster yesterday as it was toasting his bread at breakfast. Fred can now afford to move up-market and purchase a toaster that will better meet his family's needs. The aim of the MCDA will be to make the best use of the data available to inform the choice of a new toaster.

The first impact for the MCDA of these aims is on the choice of key players to participate in the analysis. A **key player** is anyone who can make a useful and significant contribution to the MCDA. Key players are chosen to represent all the important perspectives on the subject of the analysis. One important perspective is that of the final decision maker and the body to whom that person is accountable, because it is their organisation's values that must find expression in the MCDA. These people are often referred to as **stakeholders**, people who have an investment, financial or otherwise, in the consequences of any decisions taken. They may not physically participate in the MCDA, but their values should be represented by one or more key players who do participate.

Stakeholders and key players:

Members of Fred's family are the stakeholders. His wife Jane wants to consult a neighbour who recently purchased a toaster. She thinks that *Which?* magazine should be consulted too. But she and Fred don't discuss this stage. He intends to ask the advice of his local store whose salesperson he trusts to give impartial recommendations. As we shall see, failure to plan at this stage leads to a problem in the next stage.

No MCDA is ever limited just to the views of stakeholders. Additional key players participate because they hold knowledge and expertise about the subject matter. That includes people within the organisation, and often includes outside experts, or people with no investment in the final decision but who hold information that would assist the analysis. **Designers of the MCDA will need to consider what stakeholders and other key players should be involved, and the extent of their participation in the analysis.**

6.2.2 DESIGN THE SOCIO-TECHNICAL SYSTEM FOR CONDUCTING THE MCDA

When and how are the stakeholders and key players to contribute to the MCDA? That is the social aspect of the design. What form of MCDA is to be used, and how will it be implemented? That is the technical aspect. The two are designed together to ensure they are working in concert to achieve the aims of the MCDA. For example, an MCDA to support a major decision, such as the location of a new airport, will be comprehensive, covering many objectives and criteria, and will involve many interest groups and key players. The complexity of the model will in part dictate who is to contribute,

and views expressed by interest groups and key players will influence the complexity of the model. On the other hand, an MCDA to prioritise proposed projects within some governmental unit will involve few if any outsiders and will employ a simpler form of MCDA. There is no one 'best' design. **The social and technical aspects of the system for conducting the MCDA have to be considered together.**

A typical approach to problem solving in the civil service is to hold a series of meetings punctuated in between by staff work, continuing until the task is accomplished. However, an advantage of MCDA is that the process lends itself to designs that are more cost efficient than the typical approach. There are various ways to conduct this.

One approach would be to use *facilitated workshops*. These consist of participants who might be any mix of interest groups and key players. An impartial facilitator guides the group through the relevant stages of the MCDA, carrying out much of the modelling on the spot with the help of computer programs designed for multi-criteria analysis, and with appropriate displays of the model and its results for all to see. Because participants are chosen to represent all the key perspectives on the issues, the workshops are often lively, creative sessions, with much exchange of information between participants whose areas of expertise differ. Indeed, recent research³⁷ shows that the group can produce judgements that are better than could have been achieved by individuals working separately. Three factors work together to account for this enhanced performance: impartial facilitation, a structured modelling process, and use of information technology to provide on-the-spot modelling and display of results.

An impartial facilitator focuses on process and maintains a task orientation to the work. He or she ensures that all participants are heard, protects minority points of view, attempts to understand what is going on in the group rather than to appraise or refute, attends to relationships between participants, is sensitive to the effects of group processes and intervenes to forward the work of the group.³⁸ The facilitator of the MCDA assists the groups through the various stages, eliciting relevant expertise and judgements from the participants. By working as a collective, participants often discover interconnections between areas of apparently separate expertise. Each person also sees the larger picture to which the MCDA is addressed, and this larger view can affect individual contributions as their own area is put into perspective. Information technology provides for rapid construction of the MCDA model, facilitates inputting information, and shows results immediately. Because participants see each stage of the model-building process, then are shown how the computer combines the results, the overall model building becomes transparent, contributing to participants' feeling of ownership of the model. Any decisions subsequently made by the responsible person, informed by the MCDA model results, are then understood by

³⁷ See, for example, Regan-Cirincione, P. (1994) "Improving the accuracy of group judgment: A process intervention combining group facilitation, social judgment analysis, and information technology", *Organizational Behavior and Human Decision Processes*, 58, 246–70.

³⁸ Phillips, L. D., & Phillips, M. C. (1993) "Facilitated work groups: theory and practice", *Journal of the Operational Research Society*, 44(6), 533–49.

those participating, with the result that the decisions are more readily implemented.

Design the system:

Fred neglects this stage, too, and plunges ahead, visiting his local store on the way home from work. His trusted shopkeeper recommends a Philips two-slot, two-slice toaster, which Fred buys. When he arrives home, his wife says that her friend bought a toaster with a warming rack on the top, which could be useful to warm rolls and other small items. That feature is absent on the Philips. Fred realises that his criteria, even toasting and burning risk, are perhaps too restricted, that he should have thought about the process of consulting others about the features of modern toasters. He recognises that his family are all stakeholders, so he decides to explore the features and drawbacks listed in *Which?* and engage the family in discussion before making a final choice. He returns the Philips.

Facilitated workshops might last for only a few hours for relatively straightforward decisions. For complex decisions, two or three-day workshops may be required, or even a series of workshops over a period of several months (as in the case studies of Chapter 7). With thoughtful socio-technical design, workshops can prove to be helpful for a wide range of issues involving resources from tens of thousands of pounds sterling, to hundreds of millions.

6.2.3 **CONSIDER THE CONTEXT OF THE MCDA**

What is the current situation? What goals are to be achieved? Could a different frame for the issues and problems provide a recasting of the situation that would make it easier to attain the goals? What strengths can be mobilised to achieve the goals? What weaknesses might impede progress? What opportunities exist now or may appear on the horizon to facilitate progress? What threats could create obstacles? These questions raise concerns that are broader than the aims of the MCDA, but answering them will help to provide a setting for the analysis which will affect many subsequent steps.

Describing the current situation and then being clear about the goals to be achieved establishes the discrepancy between **now** and the **vision for the future** which will clarify the role of the MCDA. Presumably that gap will be closed by those authorised for taking decisions, and allocating resources to help achieve the future state. In what way can the MCDA serve the decision making process? The analysis can be framed in different ways, some more directly supporting the eventual decision, and some less so. The MCDA might be structured to:

- show the decision maker the best way forward
- identify the areas of greater and lesser opportunity
- prioritise the options

- clarify the differences between the options
- help the key players to understand the situation better
- indicate the best allocation of resources to achieve the goals
- facilitate the generation of new and better options
- improve communication between parts of the organisation that are isolated, or
- any combination of the above.

Looking at strengths, weaknesses, opportunities and threats, a SWOT analysis, is particularly useful in developing options. Keeping in mind that options are intended to achieve the goals, participants can be encouraged to generate options that will build on strengths, fix weaknesses, seize opportunities and minimise threats.

Context:

Fred and his family give some thought to what they really want. The whole family of two adults and three children often eat breakfast together; perhaps they should consider two-slot four-slice toasters in addition to the two-slot two-slice models. On Sunday morning, Fred picks up fresh bagels for Sunday brunch, so an adjustable slot width might be handy. As the family discusses what they might like, Fred suddenly remembers seeing a toaster-oven in the home of an American friend on a recent visit to the United States. Perhaps they should consider that function too, for then the oven could also be used to cook single frozen meals, which would be handy for the occasions when someone comes home late and wants a hot meal.

Other aspects of context concern the larger political, economic, social and technological (PEST) environments in which the analysis is to be conducted. Scenario analysis of how key PEST features might develop in the future, and so affect the ability of the proposed options to achieve the desired future state, sometimes stimulates key players to develop options and consider objectives that would otherwise have been ignored. Scenario analysis³⁹ can also help participants to acknowledge uncertainty about the future, and thereby make assumptions about outcomes more explicit, thus directing attention at implications which may otherwise be missed.

6.2.4 IDENTIFY THE OPTIONS TO BE APPRAISED

When options are pre-specified, it is tempting to proceed as if that is the final word. Experience shows this is seldom the case. Options are not built in heaven, they are the product of human thought, and so are susceptible to biasing influences. Groups tend to develop fewer options in situations of threat, for example, than when they are facing opportunities. A common error is to attempt to analyse just one option, under the assumption that

³⁹ Schoemaker, P.J.H. (1991) 'When and how to use scenario planning: a heuristic approach with illustration', *Journal of Forecasting*, 10, 549–564.

there is no alternative. But there is always the alternative of continuing as at present, and a proper analysis should be made of that alternative, too.

Options are often formulated on a go/no-go basis. Project funding is often conducted in this way. However, there is an alternative. Bids can be solicited to specify the benefits that would be obtained at different levels of funding. Then, some bids can be rejected altogether, others can be funded at lesser levels, and others at full levels. In this way funding decisions can be made to create more value-for-money. However, this process only works if those bidding for funds have a reasonably clear idea of the main objectives, the value that those allocating the funds wish to create.

Identify options:

Fred's family find that 23 toasters are listed in the November 1995 issue of *Which?*, and they decide to narrow their considerations to the six recommended in the Best Buy Guide. The toaster oven is more than they really need, so the idea is dropped. Fred notes that one toaster not included in the Guide, a long-slot, two-slice toaster, received a 'Best' rating for evenness of toasting, so he wants to include that as a possibility, too, even though *Which?* omitted it because it has too many drawbacks.

In all cases, whether the options are given or have to be developed, those conducting the MCDA should be open to the possibility of modifying or adding to the options as the analysis progresses. A workshop of ICL staff⁴⁰ attempted to find a way forward for a project which the managers, who were under severe budget constraints, wanted to cancel, and which the developers wanted to continue because success was only a few months away. The initial MCDA showed what everyone had perceived: that the two options of cancel and continue seemed irreconcilable. After an overnight reflection, participants considered and evaluated new options, each building on the advantages of the previous one, until finally one option was agreed. It involved a joint venture and a partial sell-off, which reduced ICL's cost but actually enhanced their benefits. It was the MCDA that stimulated the creative thought processes of participants, who otherwise would have been resolved to one group losing and the other winning. MCDA can help to create win-win situations.

6.2.5 IDENTIFY CRITERIA FOR ASSESSING THE CONSEQUENCES OF EACH OPTION

Assessing options requires thought about the consequences of the options, for strictly speaking it is those consequences that are being assessed, not the options themselves. Consequences differ in many ways, and those ways that matter because they achieve objectives are referred to as criteria, or attributes. Criteria are specific, measurable objectives. They are the 'children' of higher-level 'parent' objectives, who themselves may be the children of even higher-level parent objectives. In choosing a car, you might seek

⁴⁰ Hall, P. (1986) 'Managing change and gaining corporate commitment', *ICL Technical Journal*, 7, 213–27.

to minimise cost and maximise benefits, two high-level objectives that are in conflict. Benefits might be broken down into categories of safety, performance, appearance, comfort, economy and reliability. Safety could be considered a criterion if you use the rating given by *Which?* of how well the car will protect you in a crash. Alternatively you might wish to disaggregate safety into passive safety and active safety. Passive safety might be treated as a criterion: perhaps a count of the number of features (side bars, roll-over protection, rigid body cage, etc.) would suffice, or it could in turn be further broken down.

A useful distinction is between means and end objectives. Repeatedly ask the question ‘Why do you care about that?’ and when there is no more to be said, an end objective which is ‘fundamental’ has been reached. Take the safety criterion for cars. Why do you care about passive safety? It could reduce injuries in the event of a crash. Why do you care about active safety? It increases the chances of avoiding a crash. So for that person reducing injuries and increasing chance of survival are the two fundamental objectives. Or perhaps safety could be considered a fundamental objective if the interpretation of it given in *Which?* magazine is used: how well the car will protect you in a crash.

Which of these many interpretations are susceptible to measurement? The number of passive safety features is easy to count, but it would be a means objective, not an end in itself, so it would not be a fundamental objective. Measurement is often easier for means objectives, yet it is fundamental objectives we may care about. Perhaps experts can provide informed assessments: car safety is rated in *Which? Car* on a single numerical scale. In short, deciding on criteria to incorporate in the MCDA is very much a matter of judgement, and can require some loss in the directness with which the value is expressed in order to facilitate measurement. But as the section below on scoring the options indicates, measurement can include the direct expression of preference judgements, and these may be relatively easy even though no objective measurement is possible.

Criteria express the many ways that options create value. If options are already given, then a ‘bottom-up’ way to identify criteria is to ask how the options differ from one another in ways that matter. A ‘top-down’ approach is to ask about the aim, purpose, mission or overall objectives that are to be achieved. Sometimes overall objectives are given. The DETR’s new approach to appraisal of transport investments specifies these high-level objectives for transport schemes:

- to protect and enhance the built and natural **environment**
- to improve **safety** for all travellers
- to contribute to an efficient **economy**, and to support sustainable economic growth in appropriate locations
- to promote **accessibility** to everyday facilities for all, especially those without a car, and

- to promote the **integration** of all forms of transport and land use planning, leading to a better, more efficient transport system.

These are further broken down into criteria, some of which are susceptible to numerical measurement, including monetary valuation, others to rating, and some to qualitative description only.

Whose objectives are to be incorporated into the MCDA? Objectives often reflect the core values of an organisation. In a comparison of 18 'visionary' companies with 18 merely 'excellent' companies, Collins and Porras⁴¹ found that:

'Contrary to business school doctrine, 'maximising shareholder wealth' or 'profit maximization' has not been the dominant driving force or primary objective through the history of the visionary companies. Visionary companies pursue a cluster of objectives, of which making money is only one – and not necessarily the primary one. Yes, they seek profits, but they're equally guided by a core ideology – core values and sense of purpose beyond just making money. Yet, paradoxically, the visionary companies make more money than the more purely profit-driven comparison companies.'

Core values for some of the 18 visionary companies studied by Collins and Porras include being pioneers for an aircraft manufacturer, improving the quality of life through technology and innovation for an electronics manufacturer, technical contribution for a computer manufacturer, friendly service and excellent value for a hotel chain, preserving and improving human life for a medical company, and bringing happiness to millions for an entertainment corporation. These values infuse decision making in the visionary companies, and the authors found many instances in which profits were forgone in order to uphold the values.

Collins and Porras conclude that for these visionary companies, profit is a means to more important ends. In this sense, government departments are no different from the visionary commercial organisations—both exist to create non-financial value; only the source of their funds to do this is different. Thus, **identifying criteria requires considering the underlying reasons for the organisation's existence, and the core values that the organisation serves.**

Identify criteria:

The Jones family agrees with Fred that evenness of toasting and protection from burned fingers are essential, but four other features are included in the *Which?* review. Nobody has thought of some, like a reheat setting for warming cold toast. And eight potential disadvantages are listed. This is beginning to look unnecessarily complex; after all, it is only a toaster! A brief discussion reduces the criteria to just six: price, reheat setting, warming rack, adjustable slot width, evenness of toasting, and number of disadvantages.

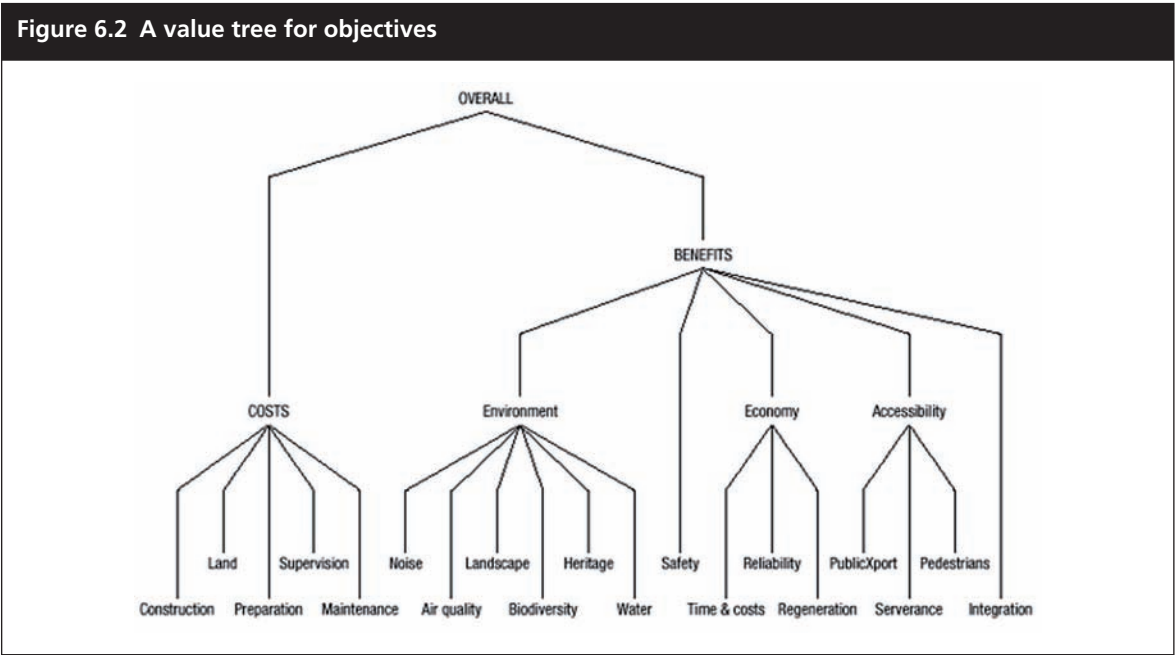
⁴¹ Collins, J. C., and Porras, J. I. (1996) *Built to Last: Successful Habits of Visionary Companies*, Century Limited, London, p.8.

The UK Treasury’s ‘Green Book’ on appraisal and cost benefit analysis states that analysis within government is concerned with effects “on the national interest.” Of course different institutions might interpret this in different ways, reflecting for example the views of experts, Ministers, senior officials, public opinion, or those directly affected by the decision. For example, the criteria used to capture distributional judgements will vary from case to case, depending upon institutional mindsets and the known preferences of the government of the day. A broadly satisfactory requirement for MCDA is that criteria should be chosen to represent the concerns of people as a whole, and to allow the expression of informed preferences.

6.2.6 ORGANISE THE CRITERIA BY CLUSTERING THEM UNDER HIGHER-LEVEL AND LOWER-LEVEL OBJECTIVES IN A HIERARCHY

Organising the criteria and objectives in this way facilitates scoring the options on the criteria and examining the overall results at the level of the objectives. The most important trade-off between the objectives appears at the top of the hierarchy. This is often between costs and benefits. Thus, the very top objective is the overall result, taking both costs and benefits into account. The next level down would show costs as one objective, and benefits as another. Costs could then be broken down into monetary costs and non-monetary costs, or short-term and long-term, or capital and operating, or any other distinction that captures more conflict between the objectives. The same applies to benefits. Top-level trade offs are not always between costs and benefits. Other possibilities include risks versus benefits, benefits to consumers versus benefits to suppliers, long-term benefits versus short-term benefits, and so forth. This hierarchical representation is often referred to as a **value tree**.

Figure 6.2 shows an illustration of how objectives and criteria for the DETR’s new approach to appraisal of transport investments might be represented:



The five objectives have been clustered under the higher-level objective 'BENEFITS', the cost of the investment has been separated out of the 'Economy' objective and represented as a separate objective, with its sub-costs represented beneath as criteria. That separation facilitates the display of benefits versus costs for schemes being appraised. There are no sub-objectives for 'Safety' and 'Integration', so those objectives also serve as criteria. This representation is meant as illustration only; it might need modification if MCDA were to be applied.

Organise the criteria:

The benefits associated with the toasters don't appear to be related to their costs, at least not for the seven toasters on the Jones's short list, so they don't bother with this step. They just want to know what is best overall.

Organising the objectives and criteria in a value tree often highlights the conflict amongst the objectives, and this can lead to refining their definitions. Making the value tree explicit and displaying it may stimulate thinking about new options that could reduce the apparent conflicts between the objectives, as in the ICL case, section 7.5. **Iterating back to previous stages is typical in any MCDA.**

6.2.7 DESCRIBE THE CONSEQUENCES

The easiest approach is to write a simple qualitative description for each option taking into account each criterion. For simpler problems, a performance matrix, as described in Chapters 4 and 5, will often suffice. For complex problems that involve a value tree, it may be necessary to construct a separate consequence table for each option, much like the Appraisal Summary Table for the DETR's new approach to appraisal for transport investments. Such a table is structured like the value tree, with separate columns (or in the case of the DETR summary table, rows) for each criterion. The bottom row usually gives the performance measures for that option on the column's criterion. Higher level objectives are shown in rows above the subsidiary criteria, throughout the table.

Describe the consequences:

The Jones family copy the data they are interested in directly from *Which?* to give the following performance matrix, which we have already met as an example in Chapter 4.

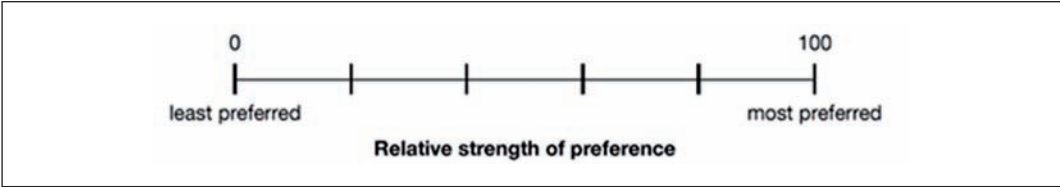
Table 6.2 Performance matrix

Options	Price	Reheat setting	Warming rack	Adjustable slot width	Evenness of toasting	Number of drawbacks
Boots 2-slice	£18				☆	3
Kenwood TT350	£27	✓	✓	✓	☆	3
Marks & Spencer 2235	£25	✓	✓		★	3
Morphy Richards Coolstyle	£22				☆	2
Philips HD4807	£22	✓			★	2
Kenwood TT825	£30				☆	2
Tefal Thick'n'Thin 8780	£20	✓		✓	★	5

A tick indicates the presence of a feature. Evenness of toasting is shown in *Which?* on a five-point scale, with a solid star representing the best toaster, and an open star the next best. The family eliminated from consideration all the toasters that scored less than best or next best.

6.2.8 SCORE THE OPTIONS ON THE CRITERIA

At this point a problem arises. It isn't possible to combine money, ticks, stars and ratings to achieve an overall evaluation of the toasters. However, apples **can** be compared with oranges, and MCDA shows how this is done. The key idea is to construct scales representing preferences for the consequences, to weight the scales for their relative importance, and then to calculate weighted averages across the preference scales. There are many ways to do all this, and the most useful approaches are described in Chapter 7.



For now, relative preference scales will be illustrated. These are simply scales anchored at their ends by the most and least preferred options on a criterion. The most preferred option is assigned a preference score of 100, and the least preferred a score of 0, much like the Celsius scale of temperature. Scores are assigned to the remaining options so that differences in the numbers represent differences in strength of preference. These are relative judgements comparing differences in consequences, and they are often easier for people to make than absolute judgements. For example, most people would agree that gaining £50,000 would give them pleasure, while losing £50,000 would cause pain, and that the pain is more undesirable than the pleasure is desirable. The pain exceeds the gain. But by how much? Is the pain twice as bad, five times as bad, 10 times as bad, or what? That is a more difficult judgement to make, but most people have some feel for it. And it is this kind of judgement that is required in MCDA. Modelling, such as that done in cost-benefit analysis, can be used for some criteria to assist in the process of converting consequences into scores that are comparable.

What do these preference scores represent? The difference-scaling method results in numbers that represent relative strength of preference. Such a measure expresses the value associated with the option’s consequence on a particular criterion. The phrase ‘strength of preference’ is here used instead of ‘value’, because the latter is often thought to imply only financial value. However, ‘strength of preference’ should not be confused with ‘preference.’ Recall that in decision theory coherent preference logically implies two measurable quantities, probabilities and utilities.⁴² Thus, A could be preferred to B if they are equal in value because A is more likely. If strength of preference is to be taken only as a measure of value, then A and B must be assumed to be equally likely. When they aren’t, then the uncertainty associated with A and B must be accommodated in some other way, as discussed later, in section 7.4, so that strength of preference measures reflect only relative value.

Score the options:

The family uses relative scaling to replace the consequences in their table with scores, with the following result:

Table 6.3 Scoring the options							
Options	Price	Reheat setting	Warming rack	Adjustable slot width	Evenness of toasting	Drawbacks	Total
Boots 2-slice	100	0	0	0	0	50	35
Kenwood TT350	25	100	100	100	0	80	61
Marks & Spencer 2235	42	100	100	0	100	50	53
Morphy Richards Coolstyle	67	0	0	0	0	100	30
Philips HD4807	67	100	0	0	100	90	49
Kenwood TT825	0	0	0	0	0	90	9
Tefal Thick’n’Thin 8780	84	100	0	100	100	0	70

Preference scores for price are, of course, the inverse of the prices; the Boots is least expensive, so scores 100 and the Kenwood TT825 the most expensive, so scores 0. All others are scaled between those limits in proportion to the inverse of their prices. For the next three criteria, if the feature is present it scores 100, otherwise 0. There are only two ratings for evenness of toasting, so the better rating scores 100, and the less good one 0. That seemed unfair at first, but later the family realises that the weighting process takes account of the small difference between the top two ratings for evenness. Finally, the family realises that the number of drawbacks can’t easily be converted to preferences because it depends on what the drawbacks are; some are more important to them than others. Instead, they simply look at the nature of the drawbacks, and directly assess preference scores to reflect their overall impression of the seriousness of the drawbacks.

⁴² Strictly speaking, utility is a measure of both subjectively judged value and the assessor’s attitude toward risk. Utilities are properly assessed using hypothetical wagers, which invoke the assessor’s risk propensities. Because the direct scaling method used here does not involve wagers, the resulting numbers are measures only of value.

Relative scaling is particularly appropriate for comparing several options presented at the same time. Sometimes, however, options are evaluated serially, so that comparison to a standard is required. It is often helpful to use fixed scales for these cases. The zero point for a fixed scale on a given criterion might be defined as the lowest value that would be acceptable – any option scoring less would be rejected outright whatever its scores on other criteria. The 100 point could be defined as the maximum feasible – this would require imagining and defining a hypothetical option as a top-scorer.

6.2.9 CHECK THE CONSISTENCY OF THE SCORES ON EACH CRITERION

This stage is usually accomplished during the process of assessing scores, but is included here separately to emphasise its importance. The method for checking consistency depends on the type of scale used. For the relative scales used in this chapter, the approach is to compare differences on a given scale. If the scale has been constructed properly, then comparing differences was a part of the scoring process, so the scale should be consistent.

In MCDA consistency of preferences is a virtue, and helps to ensure valid results. The initial assessment of scores often reveals inconsistencies, both within and between criteria. Several iterations may be needed until the key players feel that there is sufficient consistency in their preferences. The modelling process actually helps people to attain that goal; consistency is not required to start.

Check consistency:

The price is consistent with the Joneses' view that their preferences are linear (inversely) over the £18 to £30 range of prices on their list. For example, the difference in preference between the Kenwood and Marks & Spencer toasters is 42 points, the same as between the Marks & Spencer and the Tefal; both pairs differ in price by £5. Giving those two differences in price the same difference in preference means that the Jones family is indifferent between spending £5 more than £20, and £5 more than £25. If the latter increase was considered more onerous than the first, then the first decrement in preference would have been smaller than the second.

6.2.10 ASSESS WEIGHTS FOR EACH OF THE CRITERIA TO REFLECT ITS RELATIVE IMPORTANCE TO THE DECISION

The preference scales still can't be combined because a unit of preference on one does not necessarily equal a unit of preference on another. Both Fahrenheit and Celsius scales include 0 to 100 portions, but the latter covers a greater range of temperature because a Celsius degree represents nine-fifths more temperature change than a Fahrenheit degree. Equating the units of preference is formally equivalent to judging the relative importance of the scales, so with the right weighting procedure, the process is meaningful to those making the judgements.

Assign weights to the criteria:

The Joneses decide to allocate 100 points against the criteria as weights. They compare the differences between the most and least preferred toasters for all criteria, and agree that the £12 price difference matters most to them, so give it 30 points. Fred argues that he cares most about evenness of toasting, but Caroline, his elder daughter, points out that the short list is made up of toasters all of which gave above average evenness, so the difference between the top two ratings isn't very important. The family decides the small difference in evenness is about half as important as price, so it gets 15 points. The presence of a warming rack is about equal, so it receives 15 points, too. The adjustable slot seems a little less important than the price difference, but not much, so it is assigned 25 points. With just 15 points left, 10 is given to drawbacks, and 5 to the reheat feature. The ratios of those weights seems about right, though the family feels a little unsure, and Fred is still concerned about the low weight on evenness.

Most proponents of MCDA now use the method of 'swing weighting' to elicit weights for the criteria. This is based, once again, on comparisons of differences: how does the swing from 0 to 100 on one preference scale compare to the 0 to 100 swing on another scale? To make these comparisons, assessors are encouraged to take into account both the difference between the least and most preferred options, and how much they care about that difference. For example, in purchasing a car, you might consider its cost to be important in some absolute sense. However, in making the choice of a particular car, you might already have narrowed your choice to a shortlist of, say, five cars. If they only differ in price by £200, you might not care very much about price. That criterion would receive a low weight because the difference between the highest and lowest price cars is so small. If the price difference was £2,000, you might give the criterion more weight – unless you are very rich, in which case you might not care.

There is a crucial difference between measured performance and the value of that performance in a specific context. Improvements in performance may be real but not necessarily useful or much valued: an increment of additional performance may not contribute a corresponding increment in added value.

Thus, the weight on a criterion reflects both the range of difference of the options, and how much that difference matters. So it may well happen that a criterion which is widely seen as 'very important' – say safety – will have a similar or lower weight than another relatively lower priority criterion – say maintenance costs. This would happen if all the options had much the same level of safety but varied widely in maintenance costs. Any numbers can be used for the weights so long as their **ratios** consistently represent the **ratios** of the valuation of the **differences** in preferences between the top and bottom scores (whether 100 and 0 or other numbers) of the scales which are being weighted.

Implementing the swing weighting method with a group of key players can be accomplished by using a 'nominal-group technique.' First, the one criterion with the biggest swing in preference from 0 to 100 is identified. If the MCDA model includes only a few criteria, then the biggest swing can usually be found quickly with agreement from participants. With many criteria, it may be necessary to use a paired-comparison process: compare criteria two at a time for their preference swings, always retaining the one with the bigger swing to be compared to a new criterion. The one criterion emerging from this process as showing the largest swing in preference is assigned a weight of 100; it becomes the standard to which all the others are compared in a four-step process. First, any other criterion is chosen and all participants are asked to write down, without discussion, a weight that reflects their judgement of its swing in preference compared to the standard. If the criterion is judged to represent half the swing in value as the standard, for example, then it should be assigned a weight of 50. Second, participants reveal their judged weights to the group (by a show of hands, for example, against ranges of weights: 100, 90s, 80s, 70s, etc.) and the results are recorded on a flip chart as a frequency distribution. Third, participants who gave extreme weights, high and low, are asked to explain their reasons, and a general group discussion follows. Fourth, having heard the discussion, a subset of participants makes the final determination of the weight for the criterion.

Who makes up the subset? Usually the decision maker, or those representing the decision maker, or those participants (often the most senior ones) whose perspectives on the issues enable them to take a broad view, which means that they can appreciate the potential tradeoffs among the criteria. Thus, the final determination of the weights is informed by a group discussion that started from knowledge of where everybody stood, uninfluenced by others. The process also engages those closest to the accountable decision maker in making judgements that are uniquely that person's responsibility, whether or not they are expressed numerically.

The setting of weights brings to the fore the question of whose preferences count most. Chapter 2 included a discussion of whose objectives are being pursued in public sector analysis. It noted that this manual can go no further than identify this as an issue which should be recognised explicitly rather than implicitly. The choice is ultimately political and may depend on the context. However it noted that a broadly satisfactory criterion which appears to underlie many CBA valuations is that they should reflect the informed preferences of people as a whole, to the extent that these preferences and the relative importance of the criteria can be expressed in numbers. This might often be a sound aspiration for MCDA. However it may not be an aspiration which is shared, at least initially, by all those who might expect to be consulted about a particular application.

The process of deriving weights is thus fundamental to the effectiveness of an MCDA. Often they will be derived from the views of a group of people. They might reflect a face-to-face meeting of key stakeholders or people able to articulate those stakeholders' views, in which weights are derived

individually, then compared, with an opportunity for reflection and change, followed by broad consensus. If there is not a consensus, then it might be best to take two or more sets of weights forward in parallel, for agreement on choice of options can sometimes be agreed even without agreement on weights. Even if this does not lead easily to agreement, explicit awareness of the different weight sets and their consequences can facilitate the further search for acceptable compromise.

In MCDA the meaning of weights, despite these difficulties, is reasonably clear and unambiguous. The concept of a 'weight' takes on different meanings with other MCA methods used. It always needs to be handled with care.

6.2.11 CALCULATE OVERALL WEIGHTED SCORES AT EACH LEVEL IN THE HIERARCHY

This is a task for computers, though a calculator is sometimes sufficient. The overall preference score for each option is simply the weighted average of its scores on all the criteria. Letting the preference score for option i on criterion j be represented by s_{ij} and the weight for each criterion by w_j , then n criteria the overall score for each option, S_i , is given by:

$$S_i = w_1s_{i1} + w_2s_{i2} + \dots + w_ns_{in} = \sum_{j=1}^n w_js_{ij}$$

In words, multiply an option's score on a criterion by the importance weight of the criterion, do that for all the criteria, then sum the products to give the overall preference score for that option. Then repeat the process for the remaining options.

6.2.12 CALCULATE OVERALL WEIGHTED SCORES

The theory of MCDA makes clear that the simple weighted averaging calculation shown above is justified only if one particular condition is met: all the criteria must be **mutually preference independent**. This is a straightforward idea, simpler and less restrictive than real-world independence or statistical independence. It means that the preference scores assigned to all options on one criterion are unaffected by the preference scores on the other criteria. Some examples might be instructive. Two criteria can be causally linked in the real world, creating statistical correlation between the scores on the two criteria, yet be preference independent. Cars with well-appointed interiors are generally more expensive; price and poshness are positively correlated. However, most people generally prefer nicer interiors and less pricey cars. Preference scores can be given for cars' interiors without knowing what the cars cost, and for price without knowing how well appointed the interiors are. Preferences are mutually independent even though correlation exists in the real world. Take another example: choice of a main dish for an evening meal at a restaurant. In forming preferences for the main dishes on the menu, most people don't first look at the wine list.

Preferences for main dishes are independent of the wines. But preferences for wines may well depend on the main dish. Thus, preference for main dishes is independent of preference for wines, but preference for wines is not independent of preference for main dishes. Yet what goes on in the kitchen is quite independent of the wine cellar. So, a one-way non-independence of preferences exists, even though there is independence in the real world.

Incidentally, the family did not notice that the Philips toaster is in every respect at least as good as the Kenwood TT825, and on three criteria better. Thus applying the principle in Section 5.5.2.1, the Philips toaster is said to 'dominate' the Kenwood TT825, so the latter could have been eliminated from the analysis at this point.

Failure of mutual preference independence, if it hasn't been caught when the criteria are being formed, usually is discovered when scoring the options. If the assessor says that he or she can't judge the preference scores on one criterion without knowing the scores on another criterion, then preference dependence has been detected. This often happens because of double counting (see section 5.4.4.5); if two criteria really mean the same thing, but have been described in a way that apparently is different, then when the scores are elicited the assessor will often refer back to the first criterion when assessing the second. That is a signal to find a way to combine the two criteria into just one that covers both meanings.

Calculate overall weighted scores:

Caroline is just learning to use a spreadsheet, so she sets up the following table and inputs the formula for calculating the total: a simple weighted average, where the scores in each row are multiplied by the column weights expressed as decimals (eg, 30 as 0.30) and the products summed to give the total weighted score for each toaster.

Table 6.4 Calculating overall scores

Options	Price	Reheat setting	Warming rack	Adjustable slot width	Evenness of toasting	Drawbacks	Total
Boots 2-slice	100	0	0	0	0	50	35
Kenwood TT350	25	100	100	100	0	80	61
Marks & Spencer 2235	42	100	100	0	100	50	53
Morphy Richards Coolstyle	67	0	0	0	0	100	30
Philips HD4807	67	100	0	0	100	90	49
Kenwood TT825	0	0	0	0	0	90	9
Tefal Thick'n'Thin 8780	84	100	0	100	100	0	70
Weights	30	5	15	25	15	10	

Sometimes mutual preference independence fails because one or more options score so poorly on a given criterion that scores on other criteria can't

compensate. For example, people who toast breads of varying thickness may feel that the low score for a toaster with fixed slot width can't be compensated for by other features associated with high scores. This may be a signal to reject all fixed-slot toasters outright. That has the advantage of restoring mutual preference independence for the remaining options. But if that can't be done, then MCDA can still accommodate the failure by using slightly more complex mathematics, usually including multiplicative elements along with the simple weighted averaging model of this section. Multiplying preference scores causes an overall low preference if either of the two numbers multiplied together is low; this aspect of the model is non-compensatory. However, for most applications in government, particularly when fixed scales are used with the lowest position defined as the minimum acceptable, value above the minimum is additive, so the simple compensatory model is adequate.

6.2.13 EXAMINE THE RESULTS: AGREE THE WAY FORWARD OR MAKE RECOMMENDATIONS

The top-level ordering of options is given by the weighted average of all the preference scores. These total scores also give an indication of how much better one option is over another. Thus, if the total scores for options A, B and C are 20, 60 and 80, the difference in overall strength of preference between A and B is twice as large as that between B and C. Another, slightly awkward, way to express this is that compared to B, A is twice as less preferred as C is more preferred.

Another useful display of overall results is to move down a level in the value tree and display the options in a two-dimensional plot to show the main trade-offs. If costs and benefits constitute the next level down, then a graph of benefits versus costs can be instructive, for it essentially shows a relative value-for-money picture. The outer surface of the plot gives the most cost-effective options. Options appearing on the outer surface are said to 'dominate' options inside because they are both more beneficial and less costly.

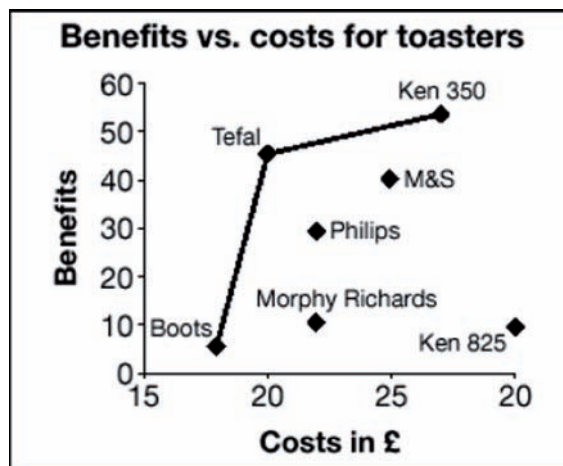
An MCDA can yield surprising results that need to be digested before decisions are taken. It may be necessary to establish a temporary decision system to deal with unexpected results and to consider the implications of new perspectives revealed by the MCDA. This temporary system consists of a series of working meetings which eventually produce recommendations to the final decision making body. At the working meetings, participants are given the task of examining the MCDA results, testing the findings for their validity, working through the possible impacts for the organisation, and formulating proposals for the way forward. When MCDA throws up surprises, it is tempting to ignore this post-MCDA stage, to demean the analysis, and find some other basis for supporting decisions. But it is important to recognise that if discrepancies between MCDA results and people's intuitions have not been explored, the MCDA model was not 'requisite'.⁴³ Exploring the discrepancies does not mean the sense of unease

⁴³ 'Requisite' is defined in section 6.2.17.

will go away; on the contrary, it could be heightened if the MCDA is found to be sound, but the message it is conveying is unpleasant or unwelcome. A period of working through the results ensures that subsequent decisions are taken with full awareness of possible consequences.

Examine results:

The family are surprised to see the Tefal as the overall winner with a total score of 70. It wasn't among the best buys *Which?* recommended! Tom, the middle child, asks if it is possible just to show the benefits separate from the costs. Caroline sets the weight on costs to zero, recalculates the benefits, then plots the overall benefits versus the costs. This shows that compared to the lower-cost Boots toaster, the Tefal provides much more benefit for little extra cost, whereas the extra benefit from the Kenwood 350 is not as cost efficient. The Tefal dominates the remaining four; it is both more beneficial and less costly. The family find the graph helps to give an overall picture of the toasters.



6.2.14 CONDUCT SENSITIVITY ANALYSIS: DO OTHER PREFERENCES OR WEIGHTS AFFECT THE OVERALL ORDERING OF THE OPTIONS?

Sensitivity analysis provides a means for examining the extent to which vagueness about the inputs or disagreements between people makes any difference to the final overall results. Especially for appraisal of schemes or projects that attract public interest, the choice of weights may be contentious. Experience shows that MCDA can help decision makers to reach more satisfactory solutions in these situations.

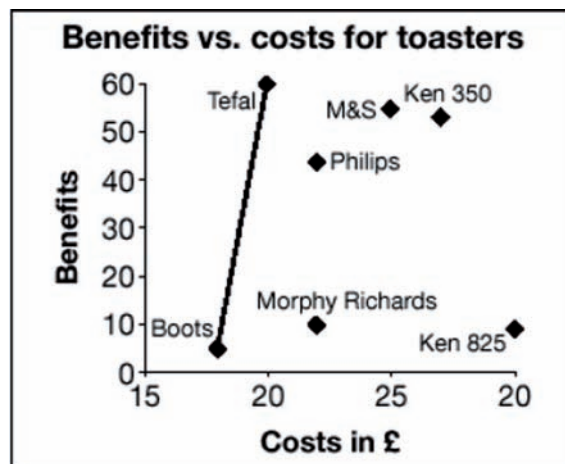
First, interest groups can be consulted to ensure that the MCDA model includes criteria that are of concern to all the stakeholders and key players. Second, interest groups often differ in their views of the relative importance of the criteria, and of some scores, though weights are often the subject of more disagreement than scores. Using the model to examine how the ranking of options might change under different scoring or weighting systems can show that two or three options always come out best, though their order may shift. If the differences between these best options under different weighting systems are small, then accepting a second-best option

can be shown to be associated with little loss of overall benefit. The reason this is usually not apparent in the ordinary thrust of debate between interest groups is that they focus on their differences, and ignore the many criteria on which they agree. Third, sensitivity analyses can begin to reveal ways in which options might be improved, as in the ICL case discussed in section 7.5.

There is a potentially useful role for sensitivity analysis in helping to resolve disagreements between interest groups.

Conduct sensitivity analyses:

Fred thinks if more weight had been given to evenness of toasting, his original purchase, the Philips, would look better overall because it received a best rating for evenness in *Which?*. Caroline doubles the weight on evenness from 15 to 30. That does indeed improve the overall benefits of the Philips, but as can be seen in the new graph, the Tefal's overall score increases as well. Fred then realises this has to be the case because the Tefal received a best rating for evenness as well, so increasing the weight on that criterion also helps the Tefal. Now the Tefal dominates all the toasters except the Boots! The family decides that their original set of weights better reflects their values.



6.2.15 LOOK AT THE ADVANTAGES AND DISADVANTAGES OF SELECTED OPTIONS, AND COMPARE PAIRS OF OPTIONS

Many analyses can be carried out to deepen understanding of the issues associated with the MCDA. These extra analyses are easily conducted with the help of computer programs designed to implement MCDA; more is said about the programs and the analyses in the chapter on results. In addition to automatically plotting graphs like those above, these programs enable users quickly to establish the advantages and disadvantages of each option, and to compare options. An advantage is a high score on a heavily weighted criterion; a high score on a relatively unimportant criterion isn't really an advantage because it doesn't contribute to overall preference. A disadvantage is a low score on an important criterion. Disadvantages are important because they reduce the overall preference, whereas low scores on unimportant criteria don't. Understanding the advantages and disadvantages helps to point to areas where options might be capable of improvement.

Comparing options is particularly useful when one option is naturally a standard. Big differences in preference scores between pairs of options on important criteria can be identified quickly, aiding the process of developing new and better options. Another helpful comparison is between the option that scores best on benefits, and the one that is least costly.

6.2.16 CREATE POSSIBLE NEW OPTIONS THAT MIGHT BE BETTER THAN THOSE ORIGINALLY CONSIDERED

The key differences between pairs of options might point to ways of generating a new option. For example, comparison of the most beneficial option with the least costly one may show how to create a new option with many, though not quite all, of the benefits of the most beneficial option, but is less costly. Sometimes this is accomplished by reducing the benefits, and thus the cost, on those criteria that do not carry much weight. Reducing the cost in this way may more than compensate for the loss of benefit, giving an option that is quite beneficial without being too costly.

If new options are generated, add them to the list of options and score the new option on all criteria. If relative scaling was used and the new option is least preferred on some criteria or most preferred on others, then it is easier to assign scores less than 0 or more than 100, respectively, so that weights do not have to be changed. An important feature of MCDA is that if the new option provides no information about the existing options and criteria, then nothing already completed has to be changed. It is only necessary to add one more preference score for each criterion, and that's all.

6.2.17 REPEAT THE ABOVE STEPS UNTIL A 'REQUISITE' MODEL IS OBTAINED

A requisite model⁴⁴ is one that is just good enough to resolve the issues at hand. Less work should be done for modest problems that are of less importance, when time is short and resources are limited. The Joneses' toaster analysis was more than requisite, although they learned some unexpected things from it. Many organisations spend unnecessary amounts of time gathering information, refining inputs and modelling. A key question to ask of any activity forming a part of an analysis is, "Will this activity, whatever its outcome, make any difference to a decision?" If not, then the activity is not worth pursuing.

An important characteristic of MCDA models is that they are often remarkably insensitive to many scores and weights. This is easily demonstrated in sensitivity analysis, but until this insensitivity has been experienced, people often find it difficult to live with rough-and-ready inputs. "Come back in six months after we have gathered more data" is a common reaction to the suggestion that an up-front MCDA will help to show what data matter, that sensitivity analysis will reveal the tolerance of final results to substantial imprecision in the many of the inputs. Many people have experience of models where precision matters. The reason that imprecision is so well tolerated in MCDA models is that the scores on many of the criteria

⁴⁴ Phillips, L. D. (1984) 'A theory of requisite decision models', *Acta Psychologica*, 56, 29–48.

will show high statistical correlation, and thus the weights on those criteria can be distributed amongst the correlated criteria in any way. In addition, changes in scores on individual criteria are often swamped by the scores for the same options on other criteria. Thus, the structure of any model that includes many criteria creates this lack of sensitivity. As experience is gained of MCDA, models become simpler and increasingly requisite.

The Joneses' toaster problem:

These last three steps weren't carried out by the family; that would have been overkill for a simple problem, and little if anything more would have been learned. So what happened? Fred and Jane decided to take a look at the Tefal and Kenwood 350 toasters. The length of the Tefal worried Jane when she saw it; *Which?* magazine hadn't given dimensions, only small photographs of the toasters. She hadn't realised how long a long-slot toaster is, and she thought that the Kenwood was rather bulky. The kitchen is small, she hates clutter on the work surfaces, and the footprint of the Kenwood seemed aesthetically more acceptable. Fred was quite taken with the warming rack as his toast often goes cold while he drinks his coffee and reads his newspaper. In the end, they decided to buy the Kenwood. The MCDA had helped them in many ways, but the final decision was theirs to take.

6.3 Uncertainty, Risk and MCDA

Because the theoretical roots of MCDA are in decision theory, it is possible to accommodate uncertainty in a coherent way. The formally correct approach is to construct a decision tree, show the consequences at the end of the tree, and then use MCDA to generate a single overall preference score for each consequence. Those scores are then folded back through the decision tree by applying the expected utility rule. This results in a probability-weighted score for each option, providing a clear overall preference ordering of the options. Unfortunately, textbooks on decision analysis treat uncertainty and multiple objectives in separate chapters, leaving the reader with the task of combining the two in real problems. Keeney and Raiffa⁴⁵ however, provide several real cases in which the two were integrated, and Hammond, Keeney and Raiffa⁴⁶ neatly blend the two approaches in a presentation that is intuitive and appealing (see their Chapters 7 and 8).

A closely related approach is useful if there is so much uncertainty about the future, and so many possible events, that the decision tree becomes unmanageable. Under those circumstances, constructing several bounding scenarios about possible developments for key external events is easier than constructing a complex decision tree. A separate MCDA is carried out assuming each scenario. This could be accomplished within one value tree

⁴⁵ Keeney, R.L., and Raiffa, H. (1976) *Decisions with Multiple Objectives: Preferences and Value Trade-offs* Wiley, New York, reprinted, Cambridge University Press, 1993.

⁴⁶ Hammond, J.S., Keeney, R.L., and Raiffa, H. (1999) *Smart Choices: a Practical Guide to Making Better Decisions* Harvard University Press, Boston, MA.

by allowing the scenarios to be represented at the highest-level branches, with exactly the same structure of objectives and criteria under each parent scenario. The scenarios can be weighted to reflect their relative plausibility or likelihood of occurring. Sensitivity analyses can be carried out on the weights given to the scenario branches to see the effects of the scenarios on the overall ordering of the options.

Uncertainty might attend only one or two criteria. If this is the case, then separate modelling of the uncertainties might allow preference scoring to be based on expected values. For example, suppose there is uncertainty about the net number of people who will experience an increase in noise if a new road scheme is built. It is usually possible to establish the minimum, most likely, and maximum number of people affected. Those three values could be used to define a triangular probability distribution that roughly describes the uncertainty. The mean of that distribution, given by the average of the three values, would then be used in the MCDA. If subsequent sensitivity analysis showed that this value could crucially affect a decision, then more sophisticated modelling of the uncertainty would be called for. This would require the careful assessment of a subjective probability distribution; see Chapter 8 of Clemen⁴⁷ for how this can be accomplished.

Another approach is to include a 'confidence' criterion in the value tree, defined as the probability that the other benefits will be obtained. Assessing a score on this criterion for a given option amounts to judging the probability that the option will create the value assessed on the other criteria. This probability is then converted to a negative penalty score which becomes disproportionately more negative the smaller the probability. Placing more or less weight on this confidence criterion provides a means of seeing how sensitive the results are to more or less concern for risk. The theory and technology of this approach are developed in Appendix 10 on Probabilities to Preferences.

Finally, some groups will wish to express the risk they feel is associated with the options, where for them risk is not just a probability or simply a reflection of uncertainty. For example, experts and the public at large may not share the same view of what risk means to them.⁴⁸ Again, it might be possible to establish a 'confidence' criterion, but options are then assessed for their relative risk, however the key players wish to define risk, using preference scores rather than probabilities.

Risk and uncertainty can be difficult topics, and there are many ways to take them into account in any MCDA. Professional help may be needed to find the best way of accommodating these concerns.

⁴⁷ Clemen, R.T. (1996) *Making Hard Decisions, an Introduction to Decision Analysis* second edition, Duxbury Press Belmont, CA.

⁴⁸ A good discussion of risk from the perspective of social scientists can be found in Pidgeon, N., Hood, C., Jones, D., Jones, B., & Gibson, R. (1992). "Risk perception", *Risk: Analysis, Perception and Management* (pp. 89–134), The Royal Society, London.

Chapter 7 Case studies

7.1 The new approach to appraisal for transport

"We are developing a new approach to the appraisal of different solutions to transport problems. This is designed to draw together the large amount of information collected as part of the appraisal of a transport problem and alternative solutions. This information is set against the five criteria which we have adopted for the review of trunk roads i.e. integration, safety, economy, environment and accessibility. It looks at the contribution of different forms of transport in developing alternative solutions and the potential effect of the new integrated transport approach, including the scope for and effect of demand management measures. It is our intention that this approach, once finalised, will be applied to the appraisal of all transport projects, including proposals for all road schemes."⁴⁹

This statement in the July 1998 Transport White Paper introduced the 'New Approach to Appraisal'. Details have been provided on how this approach has been applied to decisions on road investment projects in the Roads Review.⁵⁰ Guidance for practitioners on how the approach should be applied to road projects and other modes of transport is also available.⁵¹

The new approach to appraisal has been developed for two purposes:

- Choosing between different options for solving the same problem; and
- Prioritising between transport proposals.

The approach includes the identification and assessment of problems, the identification of options, and the assessment of these options.

NATA enables decision makers to consider the economic, environmental and social impacts of transport projects or policies in light of the Government's five main objectives (i.e. criteria) for transport:

- To protect and enhance the natural and built **environment**
- To improve **safety** for all travellers
- To contribute to an efficient **economy**, and to support sustainable economic growth in appropriate locations
- To promote **accessibility** to everyday facilities for all, especially those without a car; and

⁴⁹ DETR (1998) *A New Deal for Transport: Better for Everyone* (The Government's White Paper on the Future of Transport) The Stationery Office, p 133.

⁵⁰ See DETR (1998b) *A New Deal for Trunk Roads in England: Understanding the New Approach to Appraisal*, DETR, London.

⁵¹ See DETR (1998c) *A New Deal for Trunk Roads in England: Guidance on the New Approach to Appraisal*, and DETR (2000) *Guidance on the Methodology for Multi-Modal Studies*, both published by DETR, London.

- To promote the **integration** of all forms of transport and land use planning, leading to a better, more efficient transport system.

In turn, some of these objectives are divided into sub-objectives. In the roads application of NATA, the following three objectives are divided into sub-criteria:

- **Environment** is divided into noise, local air quality, landscape, biodiversity, heritage, and water, while the impact on carbon dioxide emissions is also distinguished separately.
- **Economy** is divided into journey times plus vehicle operating costs (VOC), highway construction and maintenance cost, reliability, and regeneration; and
- **Accessibility** is divided into public transport, severance, and pedestrians and others.

Performance against these criteria is measured where possible using established techniques, such as the time valuation methods incorporated in the COBA cost-benefit analysis programme. There are three broad types of measurement for the different criteria or sub-criteria:

- **Monetary**: where monetary values can be derived, they are used. These values are based on CBA principles
- **Quantitative**: where monetary values cannot be derived, but impacts can be quantified in non-monetary units, these are used, and
- **Qualitative**: where impacts cannot be quantified they are assessed on a scale (usually seven points, but not cardinal). Detailed guidelines have been produced for each of these scales.

An Appraisal Summary Table (AST), an example of the type of performance matrix described in Chapter 5, summarises the expected impacts of each option on a single page, where performance against criteria and sub-criteria is set out in a consistent way. As the published documents note, the 'Appraisal Summary Table cannot make judgements about the relative value to be put on the criteria and so does not provide a mechanistic way of reaching a decision. It summarises the effects in each area so that decision takers have a clearer and more transparent basis on which to make those judgements.'

The AST for one particular road option, the 16.3 km long Ferrybridge to Hook Moor section of the A1, illustrates the approach. This is a scheme to upgrade the road to a three lane dual carriageway motorway (D3M).

The top two boxes summarise very briefly, respectively, the problems the option is designed to address, and other potential options available to deal with these problems.

A1 (M) Ferrybridge to Hook Moor		1996 scheme – off line 16.3km upgrade to D3M	Cost £160m	
OTHER OPTIONS		A1 in this area caters mainly for long distance traffic, including many HGVs. Public transport solutions would not cater for sufficient traffic to relieve problem. On-line widening would require substantial demolition of properties.		
CRITERIA	SUB-CRITERIA	QUALITATIVE IMPACTS	QUANTITATIVE MEASURE	ASSESSMENT
ENVIRONMENTAL IMPACT	Noise	Over 2500 properties would experience a slight increase in noise without the scheme.	No. properties experiencing: – Increase in noise 10 – Decrease in noise 680	net 670 properties win with scheme*
CO ₂ tonnes added 2000 – 5000	Local air quality	NAQS NO ₂ objective exceeded along scheme and PM ₁₀ level increases by 2µg, but no properties close.	No. properties experiencing: – improved air quality 94 – worse air quality 0	–236 PM ₁₀ * –994 NO ₂ *
	Landscape	No significant impact. Some areas to North of scheme designated by LAs as of Special Landscape Value.	–	Slight -ve
	Biodiversity	No significant direct impact. But some habitats over 0.5km to the south of Micklefield and close to Fairburn Ings SSSI affected.	–	Slight -ve
	Heritage	Benefits to the Old Bridge across the River Aire at Ferrybridge (a scheduled monument), balanced by impact on Ferrybridge Henge (also a scheduled monument). Mitigation for latter agreed.	–	Neutral
	Water	Assuming effective mitigation, risk of damage to the water environment is likely to be negligible.	–	Neutral
SAFETY	–	Accident savings cover nearly half of the costs.	Accidents 700 Deaths 60 Serious 510 Slight 590	PVB £39m 43 per cent of PVC
ECONOMY	Journey times & Vehicle Operating Costs	Maintenance delay savings of £250m.	peak 3.1 mins inter-peak 1.4 mins	PVB £300m 330 per cent of PVC
	Cost	–	–	PVC £91m
	Reliability	–	Route stress before 142 per cent after 53 per cent	Large Low rel to PVC
	Regeneration	Serves West and South Yorkshire Assisted Area and Yorkshire and Humberside ERDF Objective 2 areas.	Serves regeneration area? Development depends on scheme?	Yes –
ACCESSIBILITY	Public transport	Small number of public transport journeys on route limit potential benefit.	–	Neutral
	Severance	Significant reduction in community severance in villages of Fairburn, Brotherton and Ferrybridge (over 680 dwellings in total)	–	Large +ve
	Pedestrians and others	Little impact on pedestrians and others	–	Neutral
INTEGRATION	–	Consistent with West Yorkshire Transport Package, Leeds and Wakefield UDPs and Regional Planning Guidance.	–	Neutral
COBA	PVB £337m PVC £91m NPV £245m BCR 3.7			

The main matrix of the table lists main criteria and sub-criteria in the two left hand columns, and then expected performance in the next three columns. The first of these provides space for qualitative comments, the second for quantitative information, and the third gives a summary assessment in terms of either monetary values, quantitative indicators, or textual rankings. To facilitate comparison with earlier road schemes, the AST also shows cost-benefit results in terms of the COBA model (which measures capital and operating costs of the road, road user time savings and vehicle operating cost changes, and changes in accident costs). The COBA results are also identified in the separate box at the bottom of the page. This box records the present value of benefits (PVB), the present value of costs (PVC), the net present value ($NPV = PVB - PVC$), and the benefit-cost ratio ($BCR = PVB/PVC$).

‘The key principle underlying the new approach to appraisal is that all the significant consequences of a road investment proposal should be set out simply and concisely using the five objectives as headings. Presenting the information in this way provides decision takers with a clear and reliable basis for their decisions, without giving prominence to any one type of impact or to benefits expressed in monetary terms compared with those that cannot be monetised.’

The DETR’s guidance document notes: ‘It is sensible to keep a record of all analyses, sensitivity tests, decisions and so on during the development of a study. Such information will explain how options have been developed and refined and why decisions have been made.’ This provides an audit trail to show how the information in each AST has been compiled.

ASTs were completed for 67 schemes in the Roads Review and for Stonehenge. The tables were then presented to Ministers to inform their decisions. They were used for sifting of schemes rather than prioritisation, and were made publicly available when the Roads Review report was published. Thirty seven schemes, costing £1.4 billion, were chosen for a targeted programme of improvements.

DETR have indicated that further research might consider weighting, particularly in connection with decisions where Ministers are not involved. This relates to the need to establish a decision rule for smaller scale investments that can be applied without reference to Ministers in each case.

Evaluation Remarks on NATA

NATA relates to an area where government decision-making has proved to be difficult, particularly because of the need to meet a number of objectives in transport policy, with actions that often have conflicting impacts on a wide range of affected parties within society.

Although transport is the area where cost-benefit analysis has been most extensively applied, it also reveals the limitations of CBA. In particular, despite the extensive work in valuing environmental effects, there has so far been little consensus among decision-makers as to the values to use; and distributional impacts have become increasingly important. The consequent

need to use more comprehensive appraisal procedures has been recognised for a long time, and particularly since the Leitch Committee recommended use of a framework approach in 1977.

The introduction of NATA represents an attempt to improve the clarity of the decision-making process. In the context of Figure 5.1, it deals with the first four, and the seventh, steps in a multi-criteria analysis. In doing so it takes a long-established decision context, that of determining priorities for transport infrastructure investment, and appraises options largely identified, as in the past, by transport engineers. NATA then clearly identifies objectives and measurable criteria to assess these options' expected performance. The assessment is undertaken by means of the Assessment Summary Tables, in effect a set of performance matrices. Decisions are then made (step 7) on the basis of the information summarised in these ASTs.

In the terms of section 5.4.4, the NATA choice of criteria appear to satisfy **completeness**, avoid **redundancy** and **mutual independence of preferences**, and reduce the possibility of double-counting (for example with regard to accident costs) which may have been a problem with the earlier procedures. The **size** (ie number) of criteria remain quite high, but this partly reflects the complexity of the decisions. The procedure appears to be **operational**, though this will be tested as NATA is used. Assessment of **impacts over time** still raises difficulties for those criteria not expressed in monetary, NPV, terms.

Since explicit weights are not used in NATA, there could be no prior certainty that the procedure would lead to consistent results. Nevertheless, a statistical analysis of the decisions made has shown that they largely reflect the performance of the schemes as shown in the ASTs.

We agree with the Department's decision not to apply standard weights to the individual criteria. However, we welcome the Department's intention to consider weighting systems which could well be applied, as described in this manual, to individual cases. This will be particularly important for delegated decision-making. As the discussion of MCA procedures in this manual has indicated (see sections 5.5.1 and 5.6) the introduction of weighting systems will also mean that explicit scoring systems will be needed for all the individual non-quantitative criteria. (Without explicit weighting, the provision of an audit trail is complex, since it requires detailed notes of reasons for decisions and information consulted.)

7.2 UK applications of MCDA

A key motivation for using MCDA to assist government decision making is the need to accommodate in formal analysis criteria that are not easily expressed in monetary terms, or which would be misleading to decision makers if monetised. In these circumstances, application of MCDA may be relatively straightforward, particularly when options and criteria are easily determined, and measures of performance exist for all the options on all

the criteria, as in the previous chapter's toaster example. However, in other circumstances vagueness and uncertainty may make it difficult to see just how MCDA should be conducted.

It may be tempting to wait, to collect additional data and clarify the situation before starting any analysis at all. However, experience suggests that an MCDA conducted early in these more difficult circumstances will guide and focus subsequent collection of data, greatly reducing the collection of masses of data that will make no difference to any decisions. How is this possible, when the MCDA model requires clarity about options, criteria and performance measures? The answer is by conducting the analysis in an iterative and reflexive fashion. The model can both shape thinking and be shaped by thinking in an evolving social process that involves all the key players. Participants use the available information, knowledge and judgements to create an initial model that helps them to understand the situation better and points to areas that could be affected by additional information. As that information becomes available, further refinements are made to the model, intuition and understanding deepens, and additional changes are made to the model. Eventually this process settles down, intuition and model results become aligned, and the way forward becomes clear despite remaining ambiguities and uncertainties. The MCDA model both models and creates a shared understanding of the way forward.⁵²

The following three case studies illustrate this iterative and reflexive process. The first is an evaluation conducted in 1994–5 by the National Audit Office of Overseas Trade Services provided by the Department of Trade and Industry.⁵³ The second is an appraisal for United Kingdom Nirex Limited of potential UK sites that could be investigated for their suitability as radioactive waste repositories.⁵⁴ The third shows how a new unitary local authority used MCDA modelling, along with the decision conferencing process, to develop a three year strategic plan for the management of their social care budget.

7.3 The NAO case study: evaluation of overseas trade services

In the autumn of 1994 the National Audit Office engaged consultants⁵⁵ to assist in the task of conducting a comparative cost effectiveness analysis of the DTI's Overseas Trade Services (OTS) export services. The following description is structured around the eight steps for applying MCDA summarised in Figure 6.1.

⁵² The theory behind this approach is explained in Phillips, L.D. (1984) 'A theory of requisite decision models', *Acta Psychologica*, 56, 29–48.

⁵³ National Audit Office (1996) *Overseas Trade Services: Assistance to Exporters*, Report Number HC 293 Session 1995–96, London: HMSO.

⁵⁴ The case reported here is based on the Proof of Evidence presented by Dr L.D. Phillips in January 1997 at the Public Enquiry into a proposed Rock Characterisation Facility at Longlands Farm, Gosforth, Cumbria, Nirex Ref PE/NRX/18, DOE Ref APP/H0900/A/94/247019.

⁵⁵ Professor Jonathan Rosenhead at the London School of Economics and Dr Larry Phillips of Facilitations Limited.

7.3.1 ESTABLISH THE CONTEXT

In an initial meeting with NAO staff, the consultant facilitators learned about the portfolio of export-related services provided to UK businesses by DTI staff in London and in 13 main UK offices, and by Foreign and Commonwealth Office commercial staff in over 200 diplomatic posts overseas, under the general heading of Overseas Trade Services. The overall purpose of these services was to improve the UK's export performance by providing information, advice and support to UK businesses about export opportunities in different world regions. The aim of the NAO's evaluation of these services was to determine their impact and relative cost-effectiveness, taking account of both non-monetary and monetary criteria, and to make recommendations for improvements. Separate analyses were to be conducted of effectiveness globally and in South East Asia. The evaluation and recommendations would be reported by the Comptroller and Auditor General to the House of Commons, with the report published by HMSO.

Key players in the evaluation included staff from Overseas Trade Services who designed and operated the services, including people seconded from the business community, and staff from the NAO. Four one-day working meetings attended by the key players were held over several months in 1995 to devise and implement the evaluation model. This case study reports an abbreviated version of the global effectiveness evaluation.

7.3.2 IDENTIFY THE OPTIONS

Although 23 services were offered by Overseas Trade Services, the NAO concluded that it would be difficult to evaluate all of them adequately in a single study, so the eleven most commonly used services were chosen for the global evaluation study. These covered 95 per cent of the activity and 65 per cent of the programme costs. To avoid unnecessary detail and to simplify the displays, seven of the eleven services are reported here. The seven services were grouped under three main headings:

Awareness and Advisory Services

- Area Advisory Groups (AAGs)
- Export Promoters (EP)
- Duty Tours (DT)

Market Information Services

- Market Information Enquiries (MIE)
- Export Intelligence Service (EIS)

In-country Support

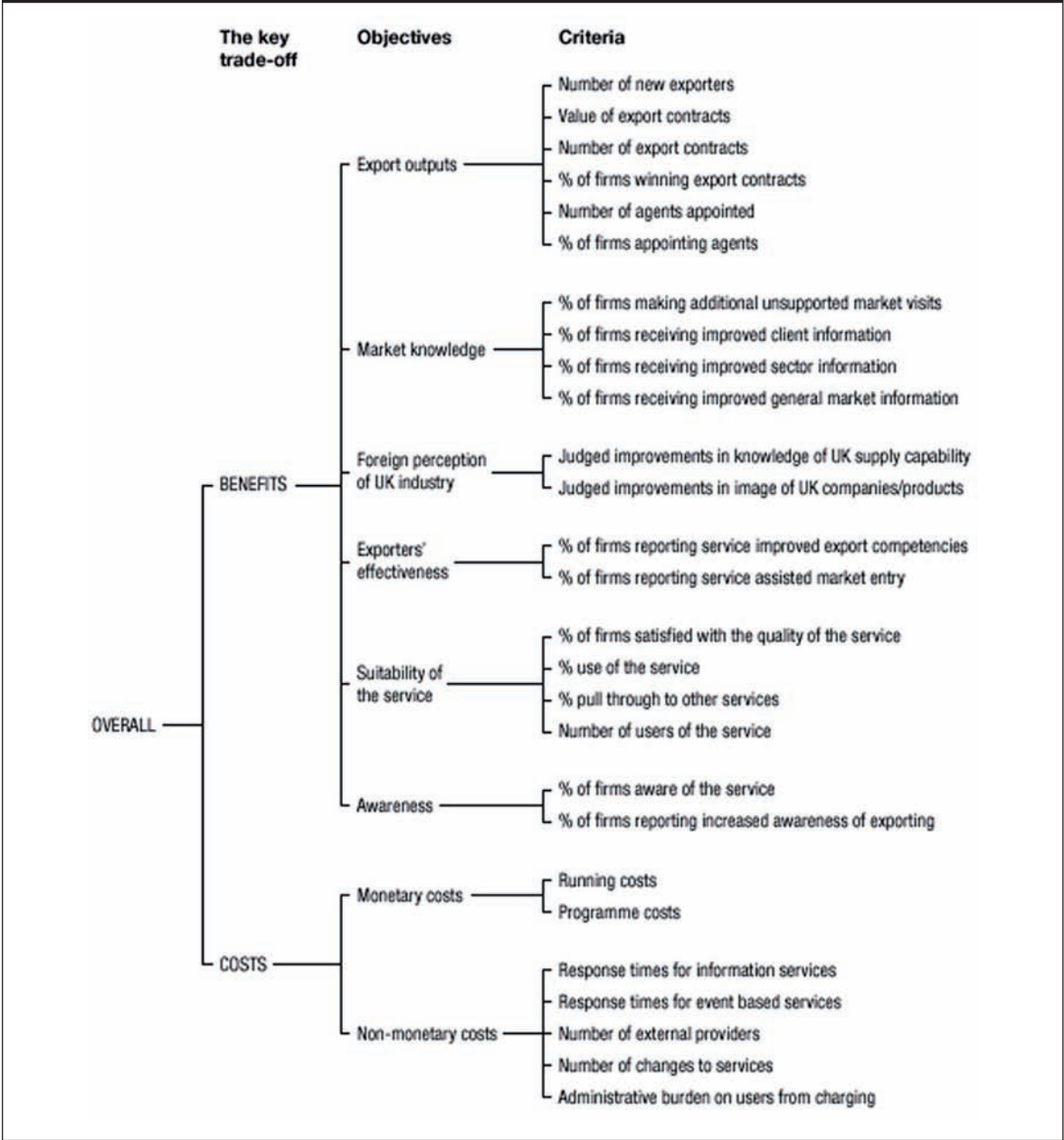
- Outward Missions (OM)
- Trade Fairs (TF)

Full descriptions of the services were obtained to ensure that all the key players understood their scope.

7.3.3 IDENTIFY THE OBJECTIVES AND CRITERIA

Objectives of government export promotion services were identified by the group in its initial meeting through discussion, consulting policy documents, and examining written descriptions of the services themselves. This enabled the group to develop a hierarchical model of objectives and criteria, a value tree, shown in Figure 7.1. (To display the value tree in portrait mode, the value tree is shown tipped on its side, with the highest level node overall shown at the left, and the bottom criterion shown at the right. Subsequent explanations will refer to higher and lower levels as if the tree is organised top-to-bottom.)

Figure 7.1 Hierarchical representation of objectives and criteria for the NAO case study



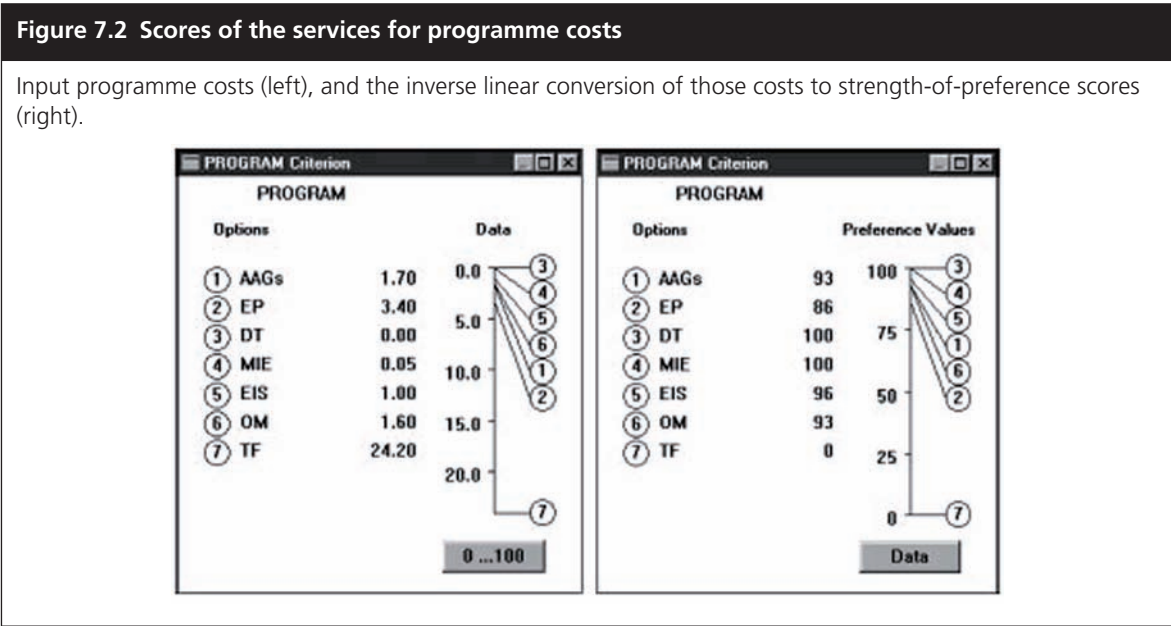
Full definitions of the criteria were agreed, along with the basis for measuring performance of the services on each criterion. For some criteria, performance was based on data held by Overseas Trade Services, and for others on the consensus judgements of the knowledgeable experts in the group, in the light of their experience with the services.

7.3.4 SCORE THE OPTIONS

The group revised the value tree at their second meeting and made an initial stab at scoring the services on the criteria. The computer program HIVIEW⁵⁶ was used throughout the modelling; at this stage it provided visual displays of the scoring of the options, and these were projected on a screen so all participants could see the displays. In the following examples performance scores are shown along with their conversion to strength-of-preference scores. In every case, the conversion used relative scaling: the least preferred service on the criterion was assigned a preference score of zero, and the most preferred service on the criterion was given a preference score of 100. All other preference scores were calculated in proportion to the inputs relative to these two anchors. HIVIEW maintains two databases, one for the input data, and one for the strength-of-preference scores.

7.3.4.1 Programme Costs

These were defined as the direct yearly costs of the services net of income per programme, expressed in millions of pounds sterling, and obtained from existing records. The left figure in Figure 7.2 shows the input performance data, the yearly costs of the services, while the right figure presents the conversion of the input data into strength of preference scores. An inverse linear function was assumed: the lower the cost, the greater the strength of preference.

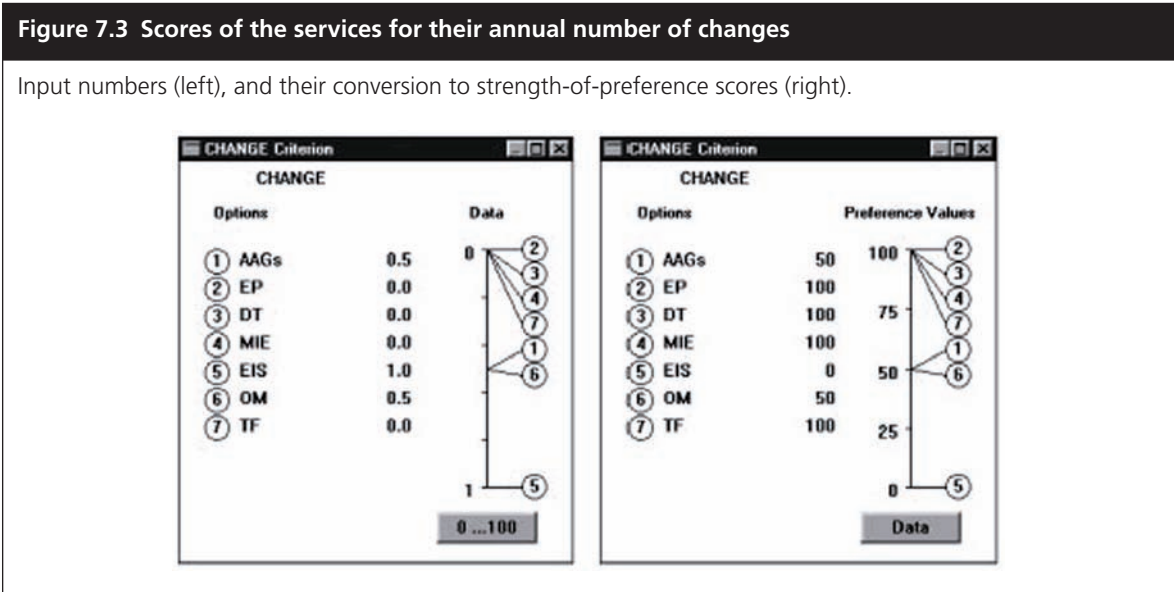


This criterion illustrates how monetary data can be used in MCDA.

⁵⁶ See the discussion of this software package in the software review.

7.3.4.2 Number of changes to the service

This was defined as the average number of times that the service delivery arrangements were changed significantly annually. The group judged that more frequent changes added to the difficulty of running the service smoothly, thereby increasing the non-monetary cost of the service. The DTI provided information over the past two years on the number of changes to each service. Figure 7.3 shows the input data and their conversion to strength-of-preference scores, again inverse linear.

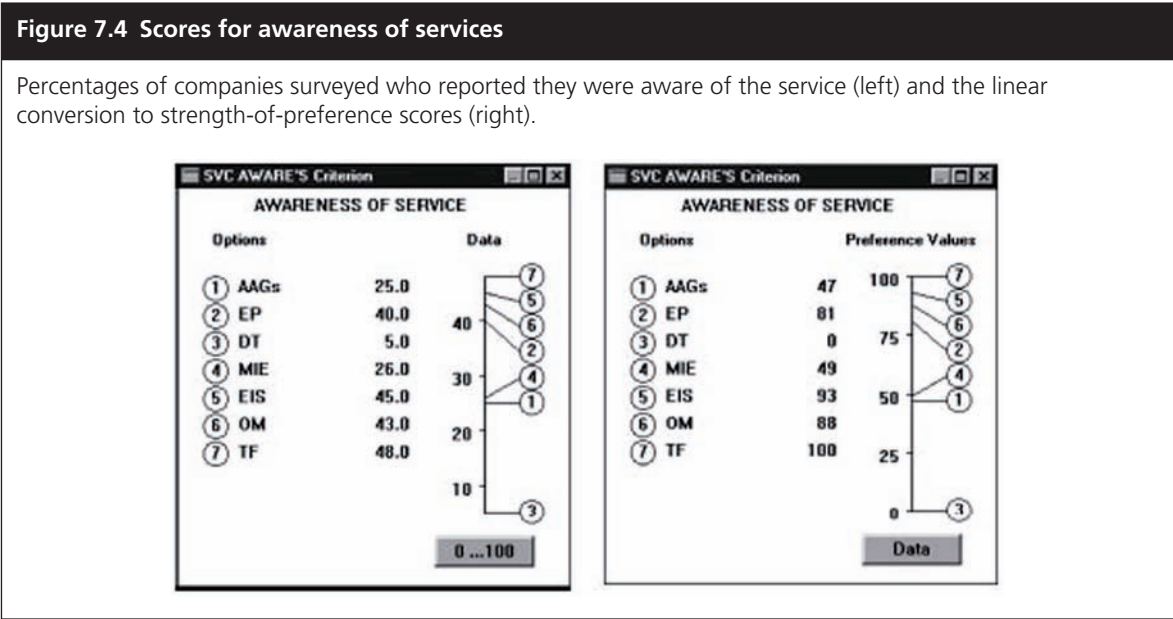


For some services no changes occurred, and for the others only one or two were made over the two-year period, so the input scale was in effect only a three-point scale. The inverse linear conversion to preferences assumes that a reduction from two to one change in two years is just as good as a reduction from one to none. This criterion shows one way of including non-monetary costs in MCDA.

7.3.4.3 Awareness of services

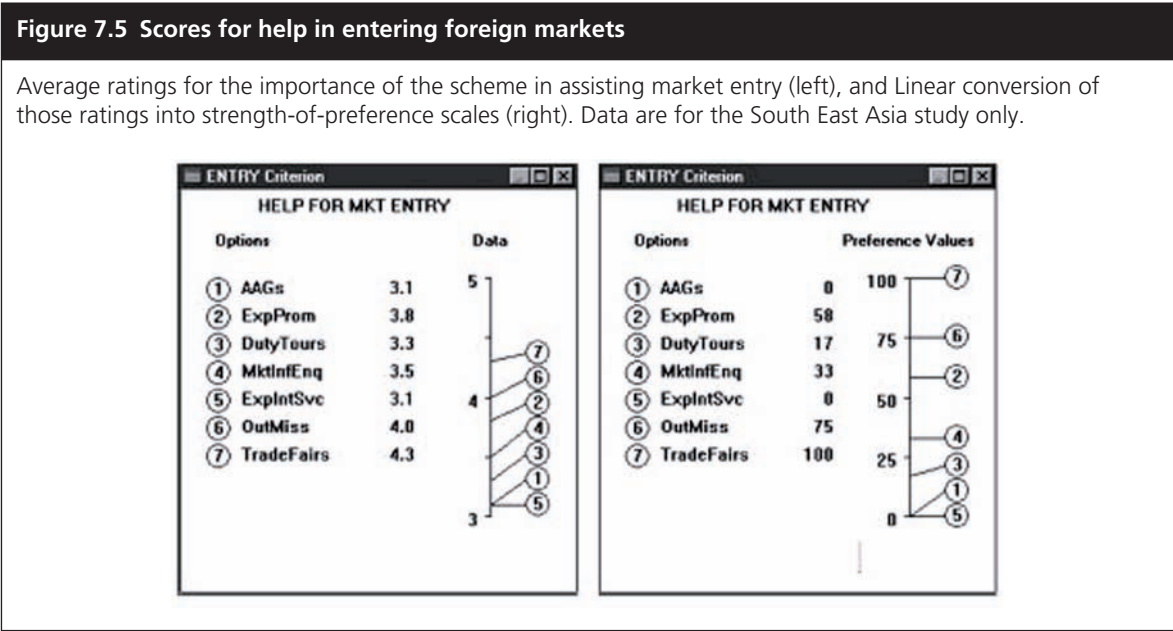
Data for this criterion were based on DTI data. The performance measure for this criterion was defined as the percentage of those businesses surveyed who reported they were aware of the service. Thus, the input data were expressed as percentages, which HIVIEW converted linearly to strength-of-preference scores. This can be seen in Figure 7.4.

This criterion shows one way to convert qualitative data (awareness of services) into quantitative data (proportion of those surveyed who reported they were aware of the service) that can be used in MCDA.



7.3.4.4 Assistance for market entry

For the South East Asia study, this criterion was defined as the average rating awarded by the 1,000 companies surveyed for the importance of the OTS scheme or service in assisting their market entry into the four South East Asian countries. Companies rated the importance of the OTS service on a 1 to 5 scale, where 1 was 'not at all important' and 5 was 'very important'. The average ratings and their conversion to preference scores is shown in Figure 7.5. Thus, this criterion shows how rating scales can be used in MCDA. Note that the conversion to preferences is linear, which assumes that the difference from one rating to the next is the same in terms of preference throughout the rating scale.

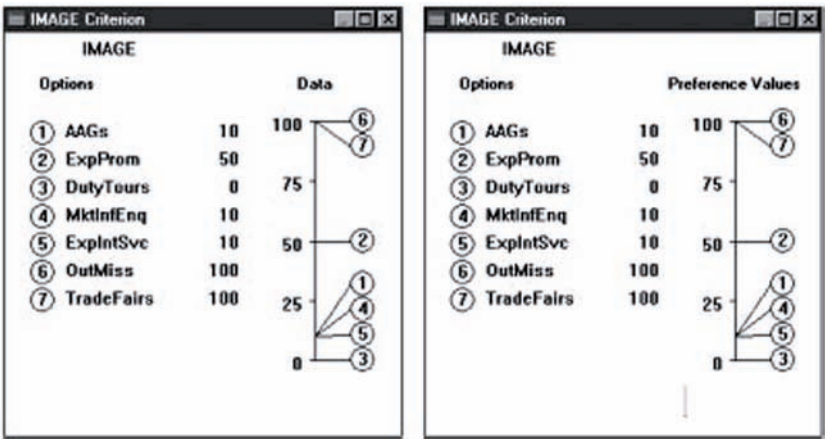


7.3.4.5 Image of UK companies/products

While tangible improvements in export performance are important, the services should also enhance the image of UK exporters. Thus, this criterion was defined as the extent to which the service improves the image of United Kingdom companies and products abroad. No quantitative data were available, so participants asked embassies and Export Promoters for their views of the relative merits of the services on this criterion. In the working meetings, participants then directly judged strength-of-preference scores as described in section 6.3.8, and checked for consistency along the lines of section 6.3.9. Thus, the two displays provided by HIVIEW are identical, as shown in Figure 7.6.

Figure 7.6 Scores for image

Directly-assessed preference scores for the extent to which the service improves the image of UK exporters and their products.



This criterion shows how directly-assessed preference scores can be incorporated in MCDA.

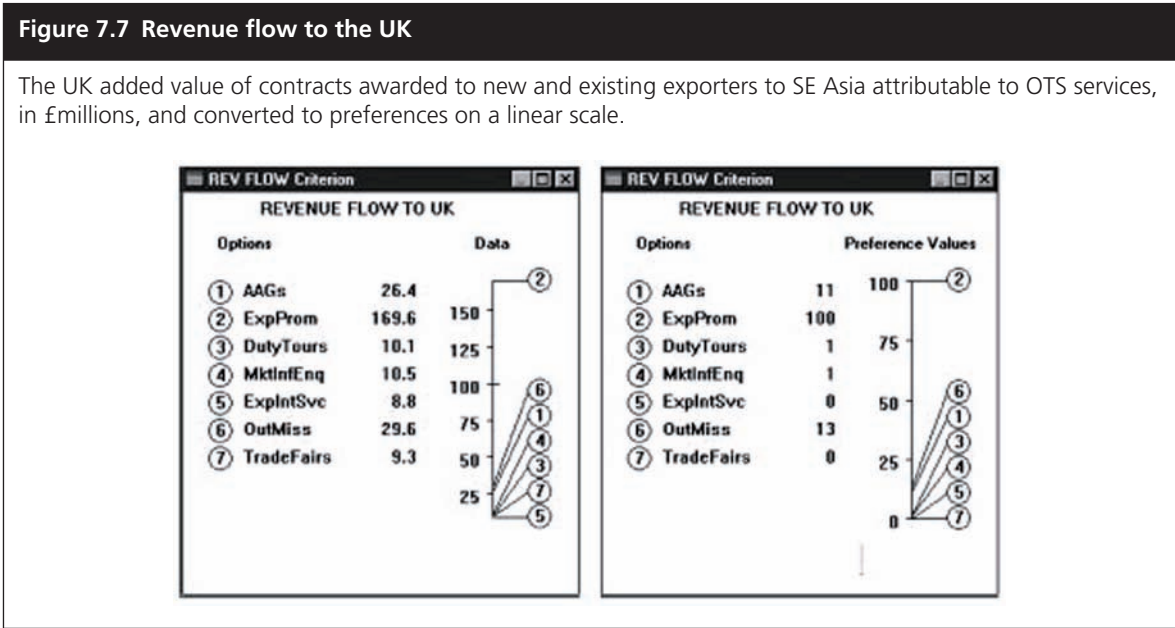
7.3.4.6 Value of export contracts

Some measure of the UK added value of contracts was required as tangible evidence that the services were affecting the flow of revenues to the UK. However, no direct measures on a company-by-company basis were available. Instead, this figure was calculated by multiplying the number of commissions by the following factors:

- the percentage of those businesses which won contracts following the use of an OTS service and who said that the service played some part in the winning of those contracts; and
- the average value of the contracts per successful survey respondent set out under the 'number of contracts' criterion.

Thus, a simple model was used to convert available data into the desired performance measure, with the result shown in Figure 7.7.

(Only data for the South East Asia study were reported in the final NAO report, so these are shown here.)



This example illustrates how performance measures derived from models can be used in MCDA.

7.3.4.7 Summary and a caution

These six examples illustrate the variety of input data that can be accommodated in MCDA: monetary and non-monetary data, qualitative data, rating scales, directly-assessed preferences, and model-derived performance measures. These examples do not exhaust the possibilities – others will be shown in the Nirex case study – but they do suggest that with careful consideration, most types of performance measures can sit comfortably within MCDA. Note that many criteria were expressed as percentages. Care must attend the use of percentages in any MCDA for if both numerator and denominator can change, preferences may be undefined. In the South East Asia export study, one thousand companies were surveyed, so the percentages always took 1,000 as the denominator. Thus, larger percentages were always preferred to smaller ones. Preference may also be undefined for a criterion that captures the percent change in some quantity when the base is different from one option to the next. In general, criteria should be operationalised with measures for which the direction of preference is unambiguous.

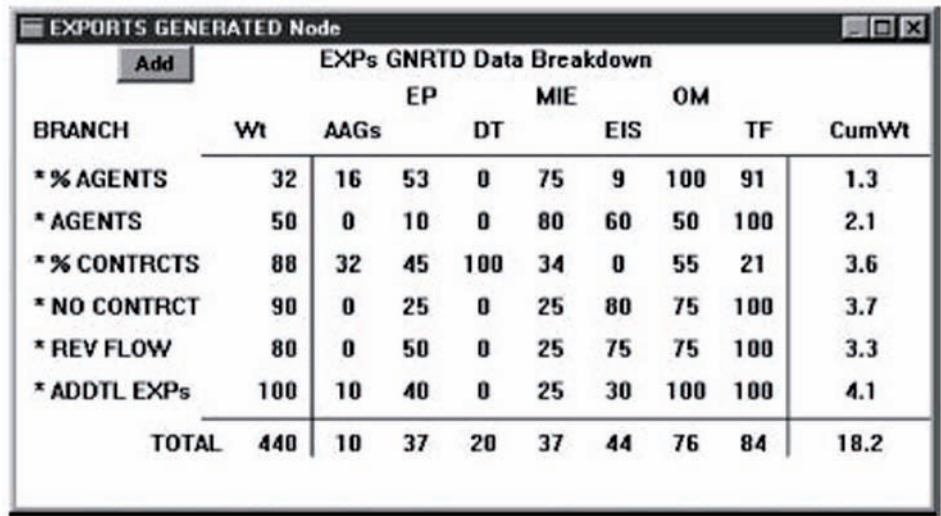
7.3.5 WEIGHT THE CRITERIA

With all input measures converted to preference scores on 0 to 100 scales, a unit of preference on one scale is not equivalent to a unit of preference on any other scale. Equating the units is accomplished by judging the relative swing in preference from the bottom to the top of one preference scale as compared to another, a method called ‘swing weighting.’ It is helpful to ask, “How different are the most and least preferred options, and how important

is that difference, on this scale as compared to that one?” As applied to the export study, paired comparisons were made of the swings on all the benefit scales to find the one scale with the largest difference that mattered. The group judged it to be the first criterion shown in Figure 7.1, the number of new exporters. Trade Fairs had generated the least additional exporters, while the Export Promoters had generated the most; this difference was judged by the group

To be the most important difference of all those on the benefit criteria. Accordingly, this criterion was assigned a weight of 100 and became the standard against which all other criteria were judged. The swings in preference on each of the other 19 benefit criteria were then compared to this one, and weights were assessed as percentages. Thus, if a swing in preference on a criterion was judged to be half that of the standard, the criterion was assigned a weight of 50. Figure 7.8 shows the weights assigned in the global study to the six criteria under the Export Outputs objective.

Figure 7.8 Preference scores for the seven services and criterion weights for the six criteria under the Export Outputs objective



EXPORTS GENERATED Node									
EXPs GNRTD Data Breakdown									
BRANCH	Wt	EP		MIE		OM		TF	CumWt
		AAGs	DT		EIS				
*% AGENTS	32	16	53	0	75	9	100	91	1.3
* AGENTS	50	0	10	0	80	60	50	100	2.1
*% CONTRCTS	88	32	45	100	34	0	55	21	3.6
* NO CONTRCT	90	0	25	0	25	80	75	100	3.7
* REV FLOW	80	0	50	0	25	75	75	100	3.3
* ADDTL EXPs	100	10	40	0	25	30	100	100	4.1
TOTAL	440	10	37	20	37	44	76	84	18.2

Note that the sum of the criterion weights is 440. Because all benefit criteria were compared to the same standard, higher level weights then become simply the sum of lower level weights. Thus, the weight placed on the Exports Generated objective is 440. By contrast, the weight on the Foreign Perception of UK Industry is just 30, the sum of the weights on the lower-level criteria of Image and Capability.

A similar weighting process was used with the cost criteria, but the higher-level consistency checks revealed anomalies in participants’ judgements of the weights. The group revised these until participants felt that the weights accurately reflected their views of the relative importance of the cost criteria.

These difficulties in weighting can arise when there are unequal numbers of criteria under one objective as compared to another. It might be, for

example, that the weight of only 30 on Foreign Perception of UK Industry compared to the much greater weight of 440 on Export Outputs was partly due to the greater number of criteria under that latter objective, six as compared to just two. On the other hand, perhaps a large number of relatively important criteria under an objective is realistic, as compared to other objectives, so the weight should be larger. By bringing these considerations to the attention of the key players, they are given an opportunity to reflect on the realism of their aggregated weights, and to change any apparent inconsistencies.

Because separate weights of 100 were assigned to most important benefit and cost criteria, and no attempt was made to compare those two criteria, it was not possible to aggregate the lower-level weights up to the level of Costs and Export Benefits. Instead, participants were invited to assess those two weights by considering how different the services were, overall, on costs and benefits. The group felt that the Export Benefits were about twice as important as the Costs, so weights of 100 and 50 were assigned at this level.

7.3.6 DERIVE AN OVERALL VALUE

The computer program does the weighted averaging arithmetic. First, all weights at each level in the hierarchy are normalised so they sum to 1.0 (but displayed as 100), and then weighted averages are calculated for each option across the criteria at a given level. As applied to the Exports Generated objective, the result is shown in Figure 7.9.

Figure 7.9 Normalised criterion weights, preference scores and totals

EXPORTS GENERATED Node

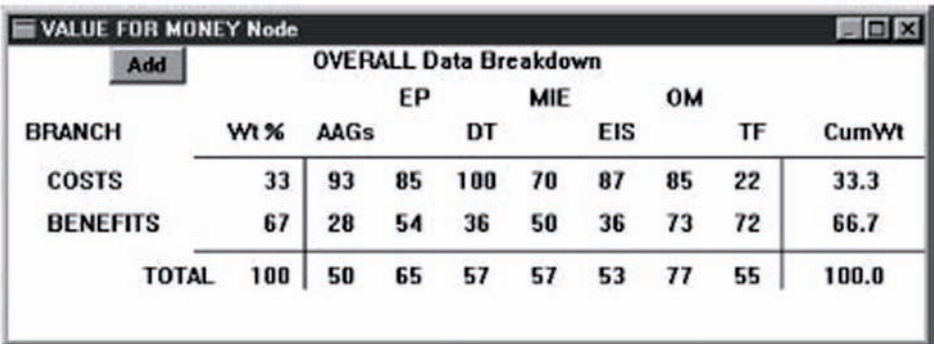
Add

EXPs GNRTD Data Breakdown

BRANCH	Wt %	EP		MIE		OM		TF	CumWt
		AAGs		DT		EIS			
*% AGENTS	7	16	53	0	75	9	100	91	1.3
* AGENTS	11	0	10	0	80	60	50	100	2.3
*% CONTRCTS	20	32	45	100	34	0	55	21	5.0
* NO CONTRCT	20	0	25	0	25	80	75	100	5.1
* REV FLOW	18	0	50	0	25	75	75	100	4.5
* ADDTL EXPs	23	10	40	0	25	30	100	100	5.7
TOTAL	100	10	37	20	37	44	76	84	24.8

Thus, for AAGs, the calculation is $(0.07 \times 16) + (0.11 \times 0) + \dots + (0.23 \times 10) = 10$. Each criterion contributes some part value toward the total of 10 at this level. Next, the computer repeats this at all levels, right up to the overall node, giving the results shown in Figure 7.10.

Figure 7.10 Overall results for the exporters study



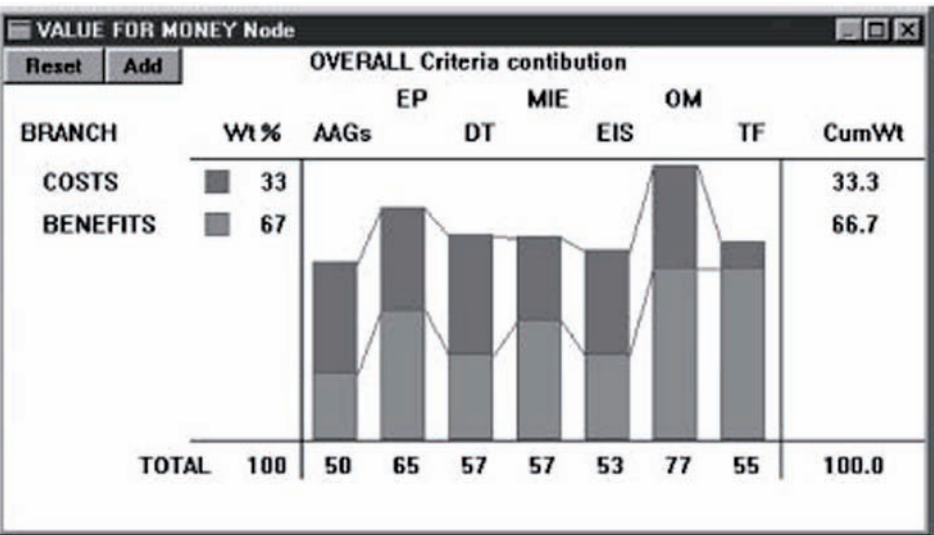
The screenshot shows a software window titled 'VALUE FOR MONEY Node' with an 'Add' button. It displays a table titled 'OVERALL Data Breakdown'. The table has columns for 'BRANCH', 'Wt %', 'AAGs', 'EP', 'DT', 'MIE', 'EIS', 'OM', 'TF', and 'CumWt'. The data is as follows:

BRANCH	Wt %	AAGs	EP	DT	MIE	EIS	OM	TF	CumWt
COSTS	33	93	85	100	70	87	85	22	33.3
BENEFITS	67	28	54	36	50	36	73	72	66.7
TOTAL	100	50	65	57	57	53	77	55	100.0

Remember that preference for costs is shown in the first row; higher preference scores mean lower costs. With export benefits judged to be twice as important as costs, the bottom row gives the overall weighted preference scores for the seven services. Outward Missions is overall most preferred, with Area Advisory Groups least preferred.

It is sometimes helpful to see the relative contribution of the lower-level scores to the total. This is shown in Figure 7.11.

Figure 7.11 The contribution of costs and benefits to the overall scores

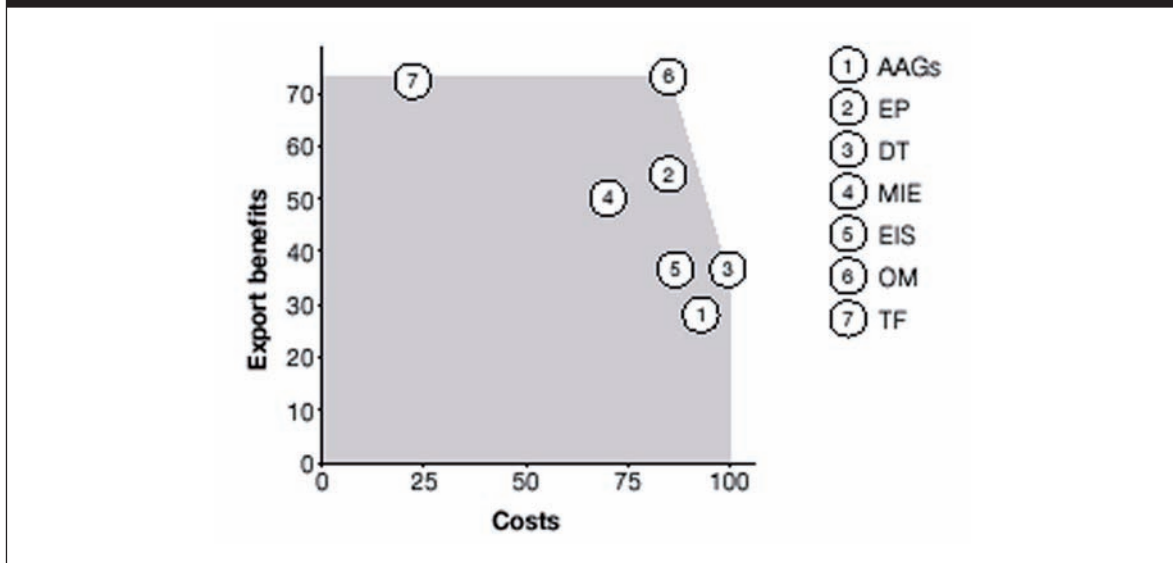


These two displays show that the Outward Missions and Trade Fairs provide substantial benefits, with Trade Fairs the least costly, while Area Advisory Groups, Duty Tours and Export Intelligence Services represent lower value-for-money.

7.3.7 EXAMINE RESULTS

These results can be shown in a manner that is more directly related to the key issue of cost effectiveness by displaying the overall benefits of the services versus their overall costs, as in Figure 7.12.

Figure 7.12 Hierarchical representation of objectives and criteria



Remember that the horizontal axis shows preference for costs, so low cost is to the right, high cost to the left. It is clear that Outward Missions is overall well positioned: low in costs and highest in benefits. All services on the outer boundary, the efficient frontier, can be considered relatively cost efficient. However, services inside the boundary are not. For example, both Outward Missions and Export Promoters dominate Market Information Enquiries at this level of the analysis; both are less costly and more beneficial than Market Information Enquiries. By examining results at one node in the value tree versus any other node, it is possible to deepen understanding of the relative merits of the options.

7.3.8 CONDUCT SENSITIVITY ANALYSIS

Very little sensitivity testing was carried out for the export study, largely because the purpose of the study, to determine relative cost effectiveness of the services, was well served by the plot of benefits versus costs. However, the sensitivity of Market Information Enquiries to inaccuracies in the data was explored to ensure that this service, which fell well inside the benefit versus cost curve, was not being penalised unfairly. The approach was to change some of the scores for this service, setting them to their most optimistic levels assuming realistic bands of error for the key data. Even when these changes were made, the service continued to fall well inside the boundary. Thus, even if more accurate data were to be obtained, Market Information Enquiries would still appear less cost efficient than the other services. This use of MCDA can help to guide data collection; it is only necessary to collect data where better data could alter the overall picture.

7.3.9 SUMMARY

When this project began, only the services themselves were clearly defined. Evaluation criteria were not yet established, and data were sparse and scattered over several locations. By the end of the first meeting, a preliminary value tree had been created; by the end of the second meeting, a revised

tree had been populated with assessments of the services based largely on participants' judgements. In successive meetings participants refined definitions of the criteria, revised the initial assessments as data became available, and conducted sensitivity analyses to see where better data might make a difference to the overall emerging picture of relative cost effectiveness.

Following the last meeting of the group, the National Audit Office updated the model as their survey data became available. The final model was not radically different from the early, largely judgement-based model, but some services changed their relative positions in the benefit versus cost plot, justifying the expense of collecting data. For example, Export Promoters appeared to be overall best in benefits, and lowest in cost, but with additional data, it moved to third place in benefits, and fourth place in costs, though it still appeared near the efficient frontier, as can be seen in Figure 7.12. Market Information Enquiries, on the other hand, remained well inside the efficient frontier throughout the study. Trade Fairs also maintained its position as high in benefits, but most costly.

7.4 The Nirex case study: appraisal of sites for further investigation as potential repositories for radioactive waste

Nirex was established to build and operate in the UK an underground repository for radioactive waste. As a first step, they screened over 500 possible sites and by mid-1988 this list had been reduced to twelve potential sites. That was too many to consider for further investigation, yet there was no obviously best site among the twelve; some sites were better on some criteria, others better on different criteria. So, in the late summer of 1988 Nirex staff approached the Decision Analysis Unit at the London School of Economics to seek assistance in reducing a set of 12 potential sites to just a few that could be recommended to the Nirex Board. An MCDA analysis had already been conducted in the United States, following a recommendation by the National Academy of Sciences, for choosing sites for further investigation, so it seemed appropriate to build on that experience in carrying out a similar analysis in the UK.

7.4.1 ESTABLISH THE CONTEXT

As was the case in the NAO case study, only limited data were available about the sites, more for some sites, less for others. After all, the aim of the MCDA was to recommend to the Nirex Board a short list of sites which would then be the subject of data gathering, so the MCDA had to proceed with limited information. Sensitivity analysis eventually played a crucial role in the analysis, for it showed which sites consistently scored well overall within the error bounds on available data.

Key players identified at an initial meeting of Nirex and LSE staff focused on the roles needed to supply information and judgements. These included people knowledgeable about repository costs, pre-closure safety and robustness, post-closure safety, transport issues, site geology, and environmental issues. Twelve people from six organisations were eventually chosen to represent these perspectives. The organisations included Nirex, UK Atomic Energy Authority, British Geological Survey, JMP (transport specialists) and Pidea (planning and environmental specialists). Five meetings, conducted as facilitated work groups and attended by these people, were held between September and November 1988. Two members of the LSE's Decision Analysis Unit facilitated the meetings and guided the MCDA modelling. All meetings were held in the LSE Pod, a specially-designed work room that facilitates interaction of participants, supports MCDA modelling on a computer, and provides easy access to technology as it is needed, but without allowing the technology to intrude. The MCDA model was implemented using HIVIEW, and the Pod's projection facilities enabled the computer's displays to be shown to the whole group.

Because the aim of the meetings was to determine a short list of sites, the Nirex Board and the relevant central government minister responsible for approving the sites were the main stakeholders. Nirex staff did not believe it would be necessary to involve outside stakeholders in the MCDA. That, they believed, would come later, when site selection itself was the issue. The final choice of a site might then involve the Treasury, local communities affected, environmental groups, and even the public at large. Also, Nirex had just completed a substantial consultation exercise about radioactive waste disposal to which a great many individuals and interest groups had contributed, and that data could be made available to the team, although it had not yet been fully analysed or reported. Nevertheless, this decision to limit the MCDA to specialists was criticised by Greenpeace at the 1995 public inquiry about Nirex's proposal to build an underground laboratory at Sellafield.

7.4.2 IDENTIFY THE OPTIONS

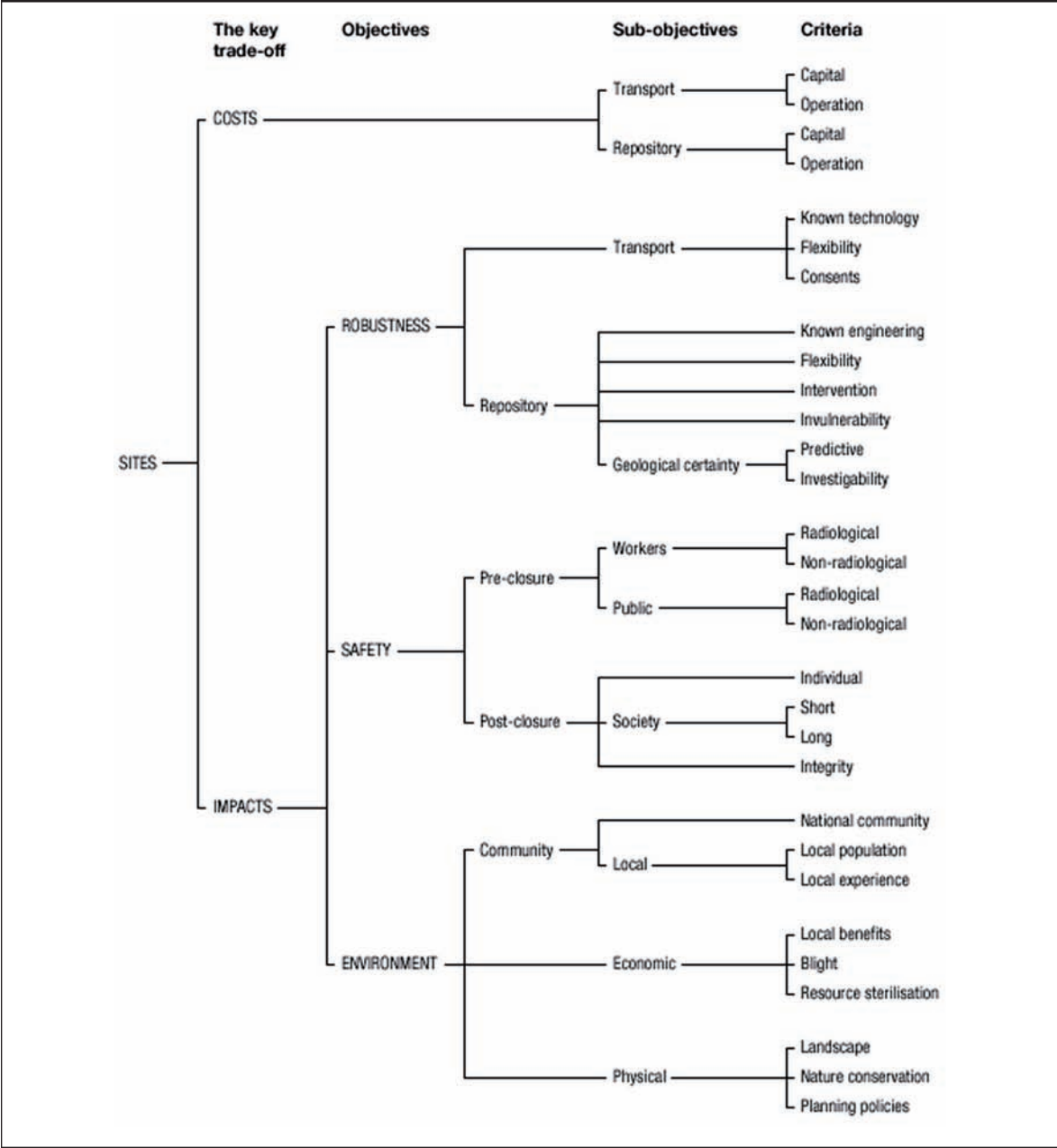
The twelve given sites were expanded at the first meeting to thirteen so a distinction could be made between two possible eastern offshore sites, illustrating that even options which are given may be revised at the early stages of the MCDA. The group was given a map of the UK, with all the potential sites identified on the map. Discussion clarified any remaining questions about the locations of the sites.

7.4.3 IDENTIFY THE OBJECTIVES AND CRITERIA

UK Government policy for radioactive waste disposal is that economic and social factors, along with safety and technical matters, should be taken into account in siting a repository. The group considered that all possible factors should be taken into account in constructing a value tree. To ensure that no key factors were omitted, the group first identified all stakeholders in the choice of a short list of sites. These included:

- Nirex Board
- Treasury
- National Environment Groups
- Regulatory Bodies
- Local Residents
- Politicians
- Local Authorities
- Scientific and Technical Community
- European neighbours

Figure 7.13 Hierarchical representation of objectives, sub-objectives and criteria for the Nirex site selection case study



The facilitators invited the group to role play any of these stakeholders and write down five factors that should be taken into account in evaluating sites. In addition, participants listed objectives that should be met by a satisfactory site, and they listed the pros and cons of the sites. Discussing these factors, objectives, and pros and cons enabled the group to construct a value tree. The higher-level objectives included:

- minimising Costs;
- ensuring the Robustness of a site, in the sense that the expected performance of a site would be confirmed as uncertainties were resolved (some sites were more predictable than others);
- maintaining a high level of Safety; and
- minimising impact on the Environment.

Each of these was broken down into lower-level objectives, and finally into performance criteria. Altogether, 30 criteria were included in the model, four under Costs, nine under Robustness, eight under Safety, and nine under Environment. These are shown in Figure 7.13.

7.4.4 SCORE THE OPTIONS

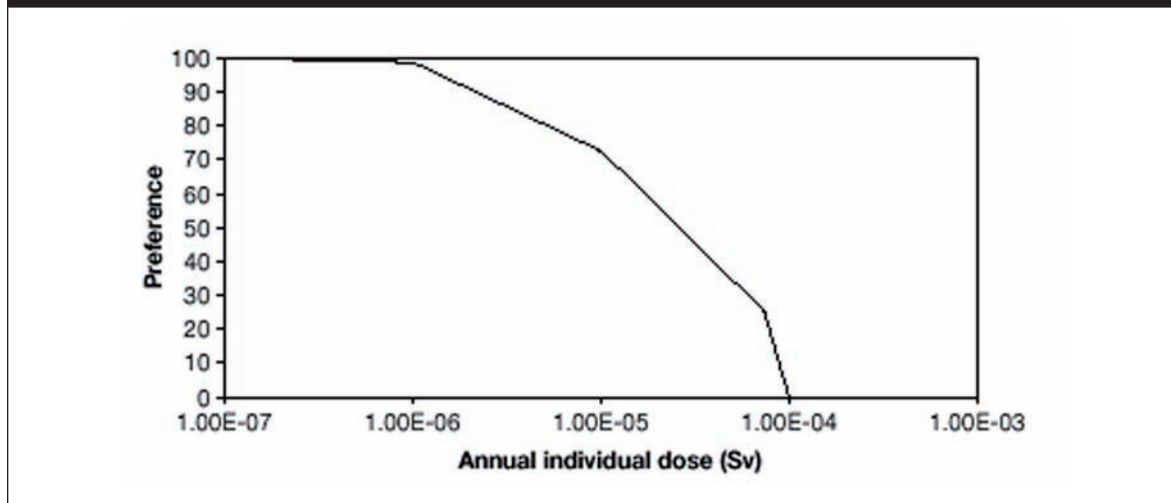
Three methods were used to obtain preference scores:

- Direct assessment. Preference scores were judged directly for some criteria.
- Rating. Rating model were constructed for some criteria, with different features of the sites gaining points that were weighted and summed to give an overall rating on an interval scale.⁵⁷ These ratings were then converted linearly to preference scores.
- Value function. A quantitative performance measure was established for some criteria, and the values of the performance measures for the site options were converted linearly to preference scores, always ensuring that the more preferred performance measures were assigned higher preference scores.

In one case the preference scores were based on a non-linear value curve. This case was post-closure safety to the individual, expressed as the annual radiation dose in Sieverts to an individual. Because considerable uncertainty attended the assessments of annual dose at most sites, participants' aversion to risk induced by this uncertainty was accommodated. To do this, a utility curve,⁵⁸ which reflects both strength of preference and risk aversion, was assessed for this criterion, but it was almost identical to the value curve, so the latter was used in the MCDA. The curve is shown in Figure 7.14.

⁵⁷ Since an interval scale has an arbitrary zero point and arbitrary unit of measurement, only ratios of differences are interpretable in terms of the quality being assessed.

⁵⁸ Assessing a utility curve requires participants to make a series of judgements comparing a sure-thing to a hypothetical wager whose two possible outcomes span the range of possible performance measures for the criterion. See, for example, Goodwin and Wright, chapter 5.

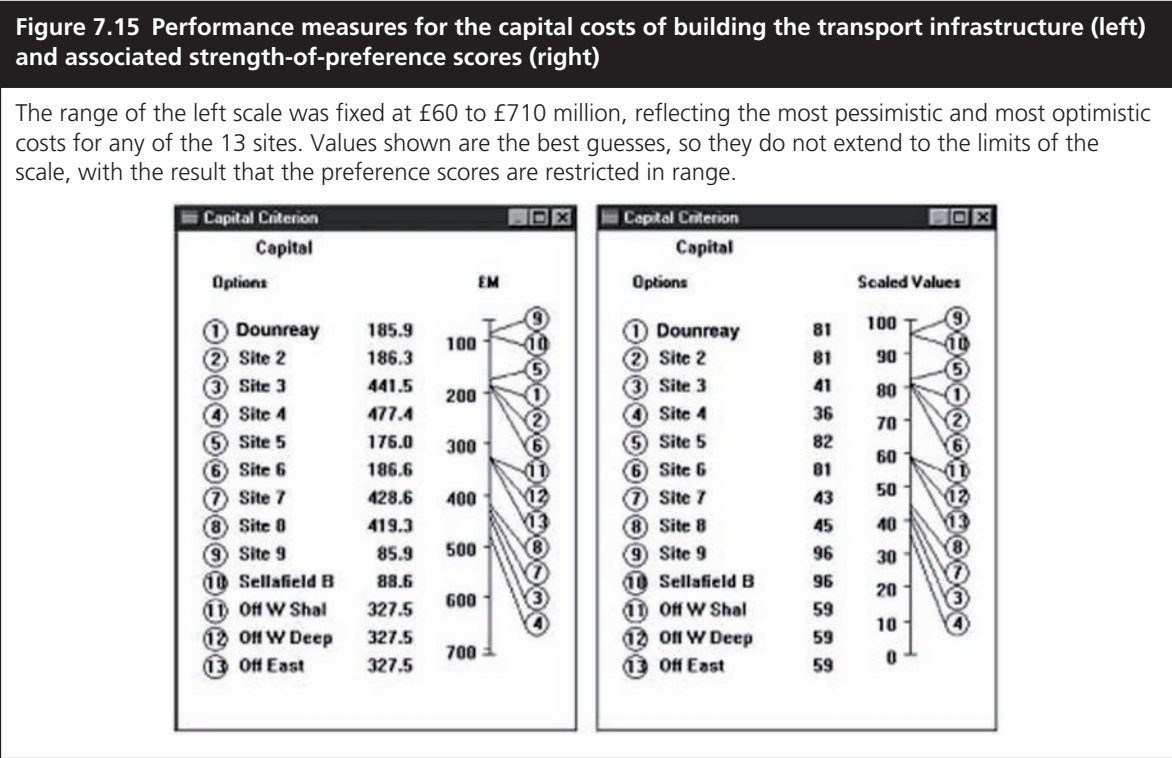
Figure 7.14 Value function for converting annual individual dose to strength-of-preference scores

Some computer programs provide for non-linear value or utility functions. It is then possible to enter performance measures for each option, and the computer uses the value function to convert those measures to preference scores. For computer programs that don't accommodate non-linear functions, such as HIVIEW, the value function is consulted off-line to convert performance measures to preferences, and then the preference scores are entered into the computer directly for each option.

Initially, the 'best guess' evaluations were used to generate preference scores of the options on all the criteria. Participants recognised that uncertainty attended many of these evaluations, so they were asked to provide 90 per cent intervals of confidence around each best guess, i.e., a low and a high value of the evaluation such that there would be a 90 per cent chance that the eventual true value would lie between those limits. Later sensitivity analyses made use of these low and high values.

By allowing performance measures to be set at pessimistic and optimistic values, the ranges on the criteria can change, and so weights have to be re-assessed to accommodate the new ranges when relative scales are used. (Recall that in relative scaling, the computer automatically sets the option associated with the least preferred performance measure on a criterion to zero and the most preferred to 100.) Computer programs differ in their ability to accommodate these range changes, so it is prudent to ensure that the program's facilities are used properly when such changes occur. One way around the problem is to use fixed scales, in which the user defines the lowest and highest performance measures to which preferences of 0 and 100, respectively, are assigned. Judicious choice of that range ensures that all subsequent performance measures will fall between the limits, thereby obviating the need to reassess weights, which apply to the fixed 0-to-100 range. An example of fixed scales is shown in Figure 7.15.

Like the NAO case study, the Nirex case shows that a variety of methods can be used in one MCDA to provide the flexibility needed to accommodate very different types of performance measures.



The scoring and subsequent weighting was accompanied by much discussion among participants in the group. An individual reporting scores that had been developed by the organisation he represented rarely went unchallenged. For example, one person might know more about costs, and another more about geology, but the geologist might challenge the assumptions about geology that were made in arriving at construction costs. By presenting their organisation’s work at developing scores for the sites on a particular criterion, each participant subjected the scores to considerable peer scrutiny, with the result that many revisions and changes were made throughout the process for nearly every score. The final performance measures, ratings and preference scores were the result of considerable discussion and debate, and represented the shared understanding of the group.

7.4.5 WEIGHT THE CRITERIA

For all the cost criteria, the weights are determined by the ranges of the costs on the fixed scales that were used on all those criteria. In comparing the first two criteria, capital and operation transport costs, the larger range of costs over the 50-year operation of the repository was for operation costs, a range of £1,650 million, so this criterion was assigned a weight of 100. The range for capital transport costs was £650 million, so this criterion was given a weight of $100 \times 650 / 1,650 = 40$. The range on the third criterion, capital repository costs, was £1,650 million, coincidentally the same as transport operation costs, so this attribute was also weighted at 100. Finally, the range on repository operation costs was £3,300 million, so its weight was put at 200. Of course the absolute values of these weights are of no importance since they are normalised before being used, but their relative values are important.

For the remaining criteria, the swing-weight method of assessing weights was used. The group recognised that different interest groups might judge the weights differently because they would care about the ranges of performance measures differently. The facilitator encouraged the group to assess the weights as appropriate to their current professional roles, and indicated that simulating different perspectives would be carried out in sensitivity analyses.

Consistency checks on the weights helped to improve their validity. For example, if criterion A had been weighted 100 and one site's preference score was 80 on that criterion, while criterion B had been assigned a weight of 80, then the group were asked if a swing from 0 to 80 on criterion A (which was weighted 100) was equivalent to the swing in preference on criterion B (which was weighted 80) from 0 to 100. In other words, a 0-to-80 swing on a criterion weighted 100 should equal a 0-to-100 swing on a criterion weighted 80. Often this sort of check was accepted, and helped the group to see they were being consistent. But when the group did not agree, they went back and revised the offending scores or weights to achieve consistency.

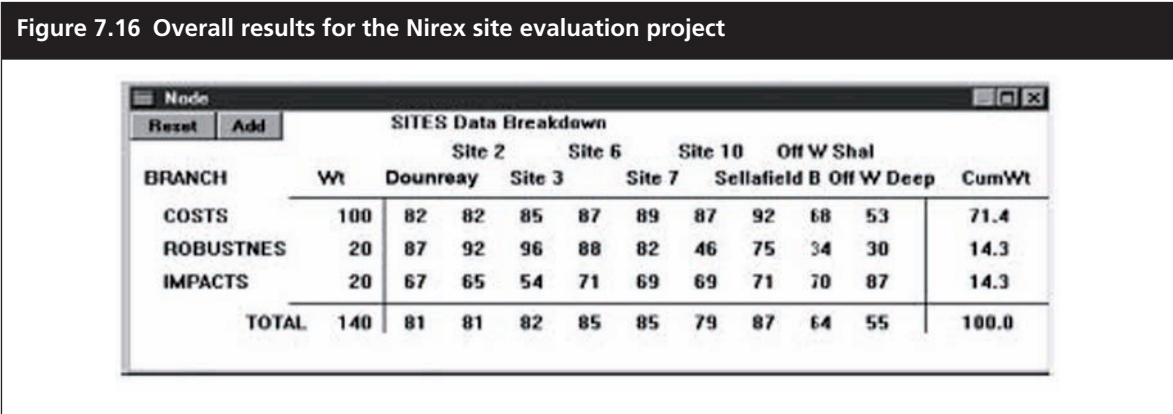
Additional consistency checks were carried out on the weights at higher-level nodes, which had been obtained by summing the weights from lower levels, as explained in the NAO case. In two cases these higher-level weights were revised. One was increased, the other decreased. Not surprisingly, only two criteria fell under the node whose weight was increased, while six criteria fell under the node whose weight was decreased, illustrating the effect of just the number of criteria under a node.

Weights on the nodes representing the four major groupings could not be made by summing lower-level weights because the criteria assigned weights of 100 within each grouping were not necessarily equal in relative importance. Of course, it would have been possible to assess relative weights for those 100-weighted scales, but instead the group elected to examine the trade-offs implied by different sets of weights on the four upper-level nodes. Equal weights gave seriously unacceptable trade-offs between costs and deaths (which occurred under some of the safety criteria). Other weighting schemes were examined, and eventually the group agreed to accept weights of 100 on costs, 20 on robustness, 10 on safety and 10 on environment, as a base case. It is worth noting that the weight of 100 on costs and 10 on safety does **not** mean that costs were considered to be ten times more important than safety. What it does mean is that the difference in costs was judged to be ten times more important than the difference in safety for the sites considered. After all, the initial sieving process from 500 to 12 sites had taken safety as a major criterion for rejecting sites, so the 12 were all considered to be potentially safe locations. The judgement of 100 on costs and 10 on safety was the consequence of valuing a life at £300,000, a figure twice that recommended at the time by the NRPB for valuing lives in a radiological protection context.

Another consistency check was obtained by ordering all 30 criteria according to their cumulative weight, the product of the normalised weights from the top of the hierarchy down each branch to the final end criterion. In a sense, the value tree can be viewed as a plumbing model: if 100 cl of liquid were poured in the top, then the amount going down each branch would be determined by the relative weights, with the cumulative weight representing the amount of liquid reaching the bottom criterion. These cumulative weights show the discriminating power of the criteria. In this case, cost criteria came out at the top because the large difference in costs between the sites mattered a great deal, while three safety criteria were at the bottom because no basis could be found to provide very much discrimination between sites already considered to be safe. It must be borne in mind that these cumulative weights were the consequences of the group’s judgements given the information available in the autumn of 1988, and may not reflect the actual differences between the sites if full and perfect information were to hand.

7.4.6 DERIVE AN OVERALL VALUE

The first, and most obvious, analysis was to look at the overall results using the base case weights. During the course of the first four meetings, four of the original thirteen sites were dropped either because they consistently scored overall very poorly whatever weights were used, or they were so similar to other sites that they could be held in reserve. Thus, the results shown here are for the remaining nine sites. Figure 7.16 gives the overall results. Note that the Impacts row combines the Safety and Environmental objectives.



The Figure shows that Sellafield B is overall the most preferred site, though several others achieve preference scores close to Sellafield’s 87. Also, Sellafield’s main advantage is Costs; a great deal of the radioactive waste is created there at Sellafield, so transport costs would be low. Other sites are overall better than Sellafield on Robustness and Impacts. Thus, weights on criteria and nodes can change the overall ordering of the sites.

To explore this effect, the group examined the overall results with different weighting systems intended to simulate different perspectives. One weighting system simulated a national environmental view: no weight

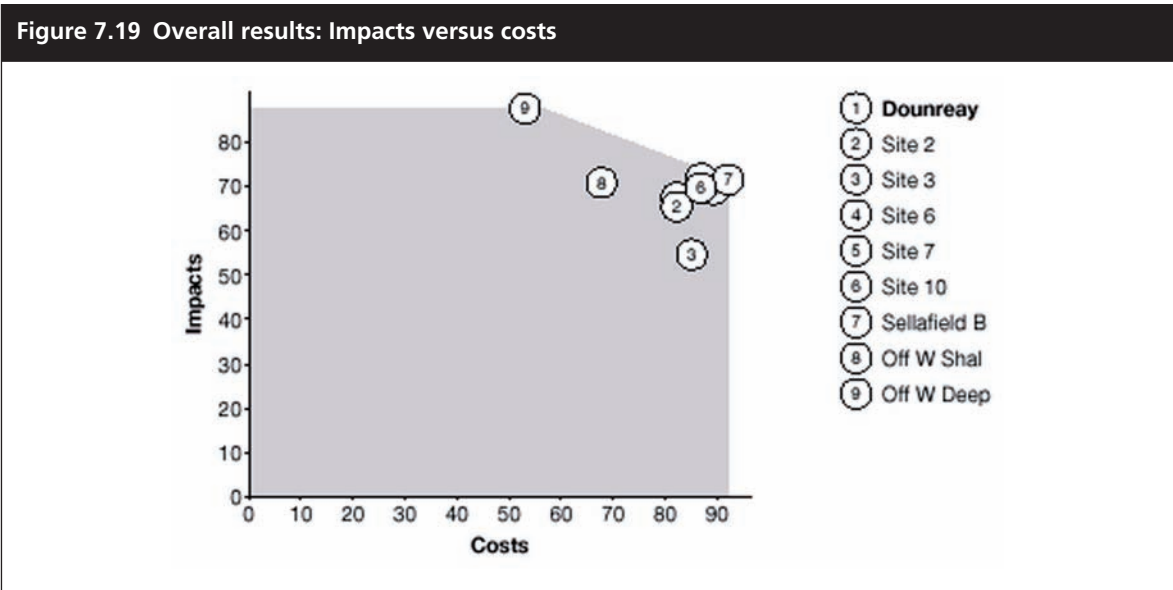
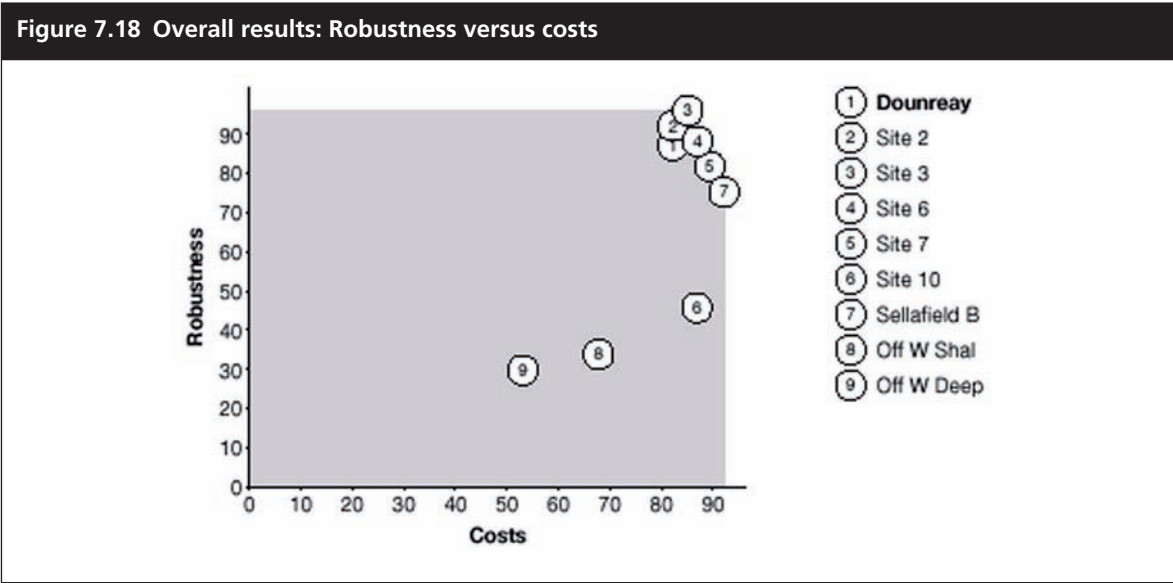
on Costs and equal weights on Robustness, Safety and Environment. An economic view was explored: a weight of 200 on Costs, 40 on Robustness, none on Safety, and 10 on Environment. Finally, a local community view was simulated: no weight on Costs, 10 on Robustness, 100 on Impact split equally between Safety and Environment, and changes to 23 lower-level weights to reflect nuances in community concerns. The results of these analyses are shown in Figure 7.17.

This table shows that the overall results are sensitive to the weights given to objectives and criteria. The Sellafield B site usually scores among the top three most preferred sites, except under the National Environmental view. Site 6 also scores in the top three, except for the Local View. No site consistently comes out most preferred.

Figure7.17 Results of sensitivity analyses on weights, simulating different perspectives					
Weights for the different perspectives:	Base	Equal	Local	National Environ'l	Economic
Costs	100	100	0	0	200
Robustness	20	100	10	20	40
Safety	10	100	50	20	0
Environment	10	100	50 & 23 others	20	10
Overall results:	Base case	Equal weights	Local view	National Environ'l	Economic view
Dounreay	81	76	60	74	82
Site 2	81	76	56	74	83
Site 3	82	72	57	68	85
Site 6	85	80	68	77	86
Site 7	85	77	66	73	86
Sellafield B	87	77	71	72	88
Offshore West Shallow	64	60	75	58	63
Offshore West Deep	55	64	83	68	50
Offshore East	16	29	58	36	15

7.4.7 EXAMINE RESULTS

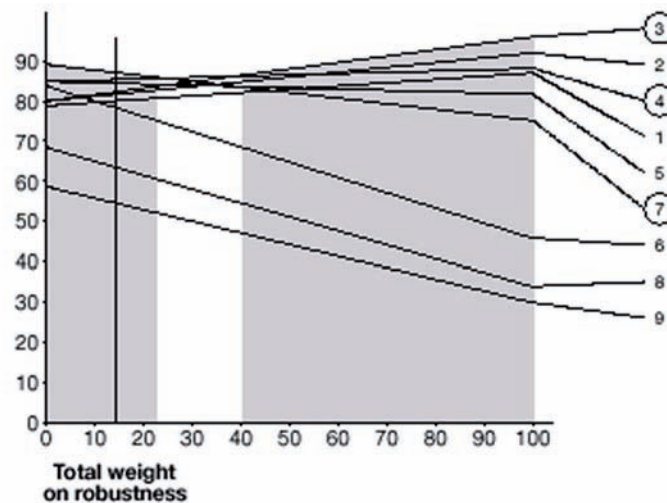
To deepen understanding of the sites, the group looked at several plots of one node versus another. Two representative examples are shown in Figures 7.18 and 7.19.



Sites on or near the efficient frontier in Figure 7.18 are 1, 2, 3, 4, 5 and 7. Figure 7.19 would eliminate 3 from that list, leaving Dounreay, Site 2, Site 6, Site 7 and Sellafield B. Site 7 is geologically very similar to Site 6, so the former could be held in reserve. This narrows the short list to Dounreay, Site 2, Site 6 and Sellafield.

7.4.8 CONDUCT SENSITIVITY ANALYSIS

Sensitivity analysis is essential and intrinsic for MCA to be useful in public policy formulation. Sensitivity analyses on individual weights provided another way of examining the model, although this material was not originally submitted by NIREX. An example is shown in Figure 7.20.

Figure 7.20 Sensitivity analysis on the weight assigned to robustness

The figure shows the effect on the overall, top-level preference score (vertical axis) of varying the weight (horizontal axis) on the Robustness node from 0 (no weight at all on this node) to 100 (all the weight on this objective). As the weight increases from zero, all other weights are reduced but kept in the same proportion to each other. The vertical line at about 14 per cent shows the base-case cumulative weight at this node. The vertical line intersects at the top of the figure with a sloping line that represents the overall preference score of option 7, Sellafield B. As the weight on the Robustness node increases, Sellafield B remains the overall most preferred option until the weight becomes about 22 per cent; then option 4, Site 6, becomes most preferred overall. With further increases in the Robustness weight to 40 per cent, Site 3 becomes most preferred overall. Sensitivity analyses of this sort were carried out for every node in the value tree and for all the bottom-level criteria. These analyses showed that over considerable variations in the weights, the sites on the short list of four, Dounreay, Site 2, Site 6 and Sellafield B were fairly consistently preferred over any reasonable range of weights. Nevertheless, critics claimed that the particular weightings settled on in deliberation by NIREX specialists, in virtually all cases, took values in the region in which Sellafield B was identified as the most preferred option. While entirely valid as subjective judgements by a group of individuals, it was claimed that this particular weighting pattern was somewhat idiosyncratic. Weighting schemes which apparently differed only marginally in a few places might yield significantly different results.

The conclusion that the four shortlisted sites were fairly consistently preferred was strengthened when pessimistic input values were substituted for the original most likely values and the model was run again under the base case weights and the national environmental view. For the latter case, the deep offshore west site moved up to overall third rank, a reasonable result because that site is very isolated. Of course, setting all input values for all the sites to their pessimistic values is not a very realistic case. A mixture of optimistic, likely and pessimistic values would be more plausible, but at the time of the analysis exploring all possible combinations was not technically feasible. Now

it would be possible to use simulation software to carry out a risk analysis, thus combining MCDA with probability modelling.

The 'sort' facility of HIVIEW made it possible to look at the relative advantages and disadvantages of each site. An advantage is a high score on a heavily weighted criterion. If the score of a site on a criterion is multiplied by the cumulative weight of the criterion, the result is the 'part score' of the site on that criterion. Summing all the part scores for a site over all 30 criteria gives the site's overall score. HIVIEW carries out this calculation and orders the criteria according to the size of the part score. It is also possible to compare two sites, using the same approach. HIVIEW subtracts the preference scores of two sites on each criterion, weights the difference by the cumulative weight, and sorts the criteria by the size of this weighted difference. This provides a display of the key advantages of one site over the other. Advantages and disadvantages of the more promising sites were checked using HIVIEW's sort facility, and all sites were compared to Sellafield B so that the group could gain a deeper understanding of the important differences between sites.

7.4.9 OUTCOME

Out of all these analyses, a picture emerged for the group of a short list of at least three sites that could be recommended to the Nirex Board. The group formulated seven recommendations. The first six explained why several of the sites should be rejected or retained. The last recommendation listed the specific sites to be investigated if 3, 4 or 5 sites were to be investigated. Each of the three short lists included Sellafield B, Site 6 **or** Site 7, and Dounreay **or** Site 2.

In January of 1989, the Secretary of State for the Environment announced that two sites would be investigated further: Sellafield B and Dounreay. In the summer of 1990, Nirex announced that it was dropping Dounreay and focusing only on Sellafield B. By 1994 enough data had been collected at the Sellafield site that Nirex proposed building an underground laboratory, a Rock Characterisation Facility, to verify *in situ* their findings to date, but Cumbria County Council refused planning permission and the proposal went to a public inquiry. This was completed early in 1995 and the Inspector upheld Cumbria's refusal. One of his criticisms was that sites other than Sellafield B should have been examined, specifically Site 6. The MCDA approach was not criticised, though the Inspector suggested that more weight should have been given to Safety, and he used the MCDA to justify the potential attractiveness of Site 6.

7.5 Medium term revenue budget planning

A unitary local authority wanted to develop a three-year strategic plan for the management of their social care budget. They engaged ICL's Local Government Consultancy service, an organisation that has specialised in applying MCDA to budget allocation in the public sector, to help them in this

process. Over 200 decision conferences have been facilitated by the ICL team for over 50 local authorities on budgeting issues. The following description was written by Martin Brader, a member of ICL's team.

7.5.1 ESTABLISH THE DECISION CONTEXT

The current grant regime (Standard Spending Assessment or SSA) implied a reduction in the budget of over 5 per cent in real terms. The following year the social care services were also to be the subject of a 'Joint Review', i.e. an external inspection. The management team decided to hold a two-day financial planning workshop using the *decision conferencing* process to help them develop their strategies.

7.5.2 IDENTIFY OBJECTIVES AND CRITERIA

Two months before the workshop a planning meeting took place between the corporate director responsible for social care and the ICL consultant who was going to facilitate the workshop. At this meeting decisions were made on:

- a) the objectives for the workshop (see h, i, j below)
- b) the workshop participants
- c) how the various service areas would be grouped together
- d) how the options were going to be prepared and presented
- e) the number of options and criteria which could realistically be assessed in the workshop
- f) asking the option preparers to identify savings totalling –15 per cent and service developments costing no more than +5 per cent, in order that there would be sufficient options available for evaluation; and,
- g) the date and venue for the workshop.

The workshop objectives were:

- h) to contribute towards the production of a 3-year strategic plan for the management of the social care budget within the current grant regime;
- i) to produce proposed priorities and budget options so that members and officers can achieve a social care spend of 10 per cent above SSA within financial year 2001/02; and,
- j) to enable the council as a whole to assess the suitability of using the decision conferencing process in other situations.

Shortly after the planning meeting the consultant facilitated a half-day meeting with the proposed workshop participants. At this meeting, the decision conferencing process was illustrated to familiarise the participants with the process to be used in the two-day financial planning workshop; the workshop participants were briefed on how to prepare the options which were going to be evaluated at the two-day workshop; and the criteria by

which the options were going to be evaluated were identified and defined (see below).

The following criteria were established:

- k) The extent to which our option improves PROTECTION for the most vulnerable
- l) The extent to which our option promotes INDEPENDENCE
- m) The extent to which our option supports CORPORATE STRATEGY and/or NATIONAL OBJECTIVES

Note that it was important to establish the criteria before the options were prepared. This enabled the option preparers to take account of the criteria when they were creating their options.

7.5.3 IDENTIFY THE OPTIONS TO BE APPRAISED

The option preparation process involved:

- a) The accountants providing service cost information
- b) Service managers and their teams identifying potential options for both savings and service developments
- c) A progress review meeting with the ICL consultant
- d) A review of the options by the senior management team to ensure that all the options were realistic and that there were no obvious omissions; and,
- e) All participants being circulated with a copy of all options several days before the workshop took place.

7.5.4 SCORE THE OPTIONS

All the scoring took place during the workshop. Seventy two options were prepared, each of which needed evaluating against each of the three criteria.

Early into the scoring it became apparent that the third criterion CORPORATE STRATEGY and/or NATIONAL OBJECTIVES was difficult to score. This was largely due to a lack of clarity about the meaning of the criterion. The group refined the definition of this criterion and renamed it DIRECTION. The scoring of each option against each criterion took place during the first day and a half of the two-day workshop. An example of the scoring is shown below:

Option	Benefits		
	PROTECT	INDEP	DIRECTION
CF11: Leaving care	0	+10	+10
CF12: Foster payments	+20	+20	+100
CF13: Family centre	+10	+15	+20

Three illustrative options are shown. The option preparers had provided a full description of each option and its impact to the participants prior to the workshop. The scoring process started with the group identifying a 'benchmark' option for each criterion that was given a score of 10. All options were then scored relative to this benchmark and also to each other. In the example above option CF11 is being used as a benchmark for criteria INDEP and DIRECTION, and CF13 the benchmark for PROTECT. It is important to note that the scores are all relative to one another **within each** criterion, eg for option CF12, the score of +20 is relative to the +10 score for CF13 under criterion PROTECT. At this stage no relationship had been established between the PROTECT scores, the INDEP scores or the DIRECTION scores.

During the scoring 11 of the options were modified (either removed or combined with other options) and three new ones were created. All scoring was completed by lunchtime of the second day.

7.5.5 WEIGHT THE CRITERIA

In order for a combined value (or benefit score) for each option to be calculated it is necessary to establish the relationship (or relative weights) between the scores in each criterion. The purpose of the Criteria weights is to enable all scores to be converted to a common scale. In the example above the criteria weights establish how much a PROTECT point is worth compared to an INDEP point compared to a DIRECTION point.

In general, groups find this a very hard judgement to make. In fact in this workshop the participants could agree that PROTECT and INDEP should carry more weight than DIRECTION, but couldn't agree whether PROTECT should carry more weight than INDEP or that they should have equal weight. Finally the group agreed to use two different sets of weights to reflect all participants' views. An initial model was built using the set of weights favoured by most participants. The second set of weights was used during the Sensitivity analysis, see 7 below.

7.5.6 DERIVE AN OVERALL VALUE

Implementation of the model was achieved using the Equity software (see SOFTWARE section at the end of this manual). With all scores and weights available, the program calculated a weighted average of the three benefit scores for each option, giving the total relative value, or benefit, of each option. Since service cost information for each option had been input to the program, each option was now characterised by two numbers: its service cost and its total benefit. Equity calculates the ratio of these two values, a benefit-to-cost ratio that is a single index of the priority of the option: larger ratios signal a higher priority option. It is helpful to visualise this ratio as the slope of the triangle shown on the left. One triangle is therefore associated with each option. Equity initially stacks the triangles in the order the options appear in the input data, giving a graph of cumulative benefit versus cumulative cost. The graph below is such a plot for the Service Area 'Children and Families.'

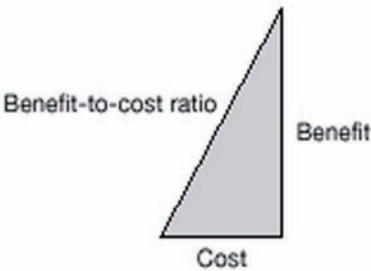
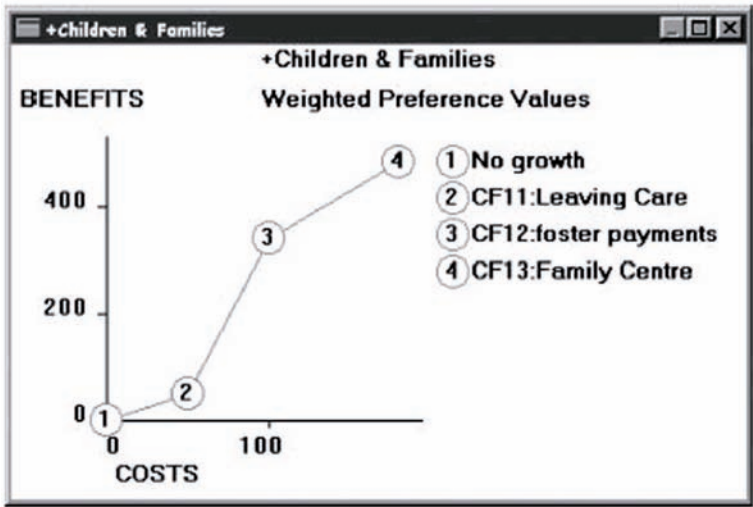
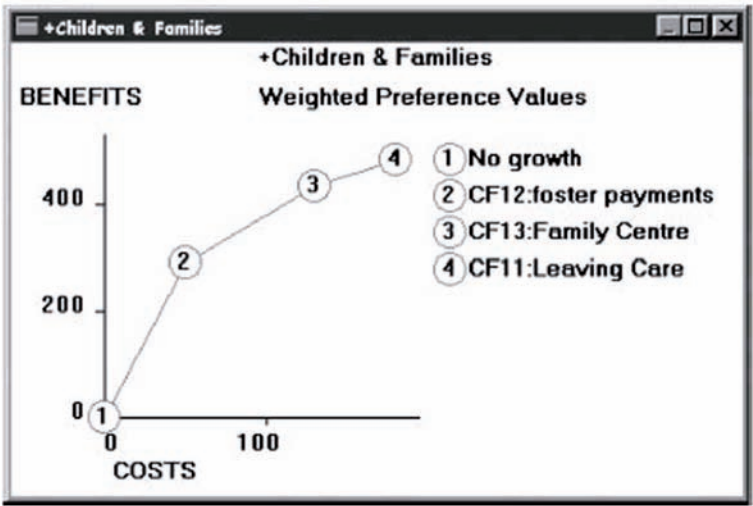


Figure 7.21 Children and families: Weighted preference values



Note that the benefit-to-cost ratio for option 2, CF11, shown as the slope of the line from option 1 to 2, is not as steep as the slope from option 2 to 3. In other words, option 3 is better value for money than option 2, so it should be shown in a higher priority position. Equity was then instructed to re-arrange the options in order of declining slope, a process called 'sorting' the options. The resulting sort gives the options in order of their benefit-to-cost ratio, as shown in the graph below.

Figure 7.22 Children and families: Sorted results



This shows that the option expected to deliver most value for money is CF12, followed by CF13 and finally by CF11. This sorting of options is important because the budget constraints may restrict funding of this area, preventing all three options from being chosen. Equity always chooses options in whatever order is shown, so it is obviously better to arrange the options in a value-for-money priority. The priorities are seldom obvious when several benefit criteria are considered, so the sort process helps to identify the higher priority options. Decision makers often think mainly about the benefits to be achieved, not the ratio of benefits to costs. They are often surprised when high-benefit projects turn out to be low priority because they are very costly, so poor value-for-money 'sacred cow' projects often exhibit this characteristic. Low value options can achieve high priority if they are also low in cost. Several low-cost, low-benefit items together can provide substantial benefit that was completely hidden because nobody thought of the options as forming a collectively good buy. These options are sometimes characterised as 'sleepers,' or 'low-hanging fruit.'

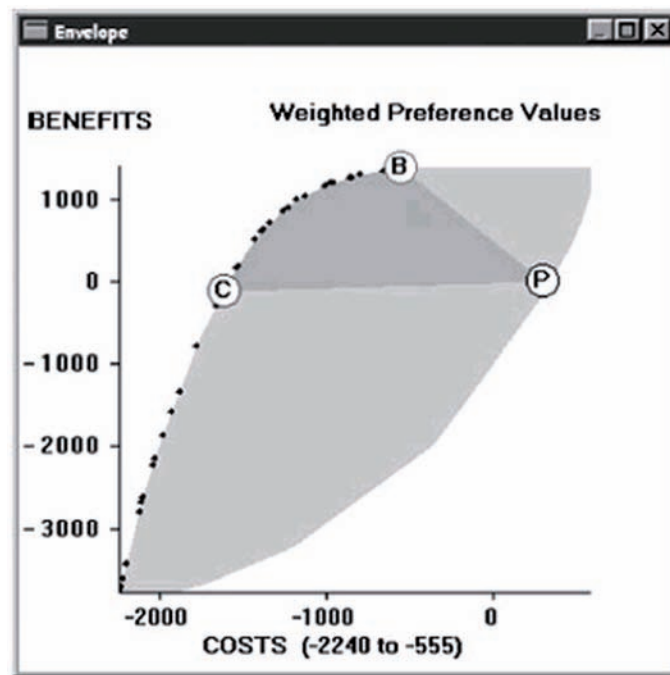
7.5.7 EXAMINE THE RESULTS

The analysis continued by using the set of weights favoured by most participants, and initially considered the value for money of the current allocation of resources.

There were approximately 32 million alternative budgets (or ways of combining the options) within the model. Each of these is associated with a single total cost, obtained by summing the costs of the individual options, and a single total benefit, also the sum of the individual benefits. All these budgets appear within the light or dark shaded areas of the figure below. The shaded areas define an envelope; there are no low cost, high benefit budgets (upper left), nor are there any high cost, low benefit budgets (lower right). The points on the upper surface show the locations of the best budgets for a given total cost. Each point can be decoded in Equity to show how the limited resources are allocated to options across all areas. The 'P' (Present Position) represents the current allocation of resources.

The Equity software identified and labelled two additional points, B and C. Point C was obtained by moving horizontally to the left to find a budget whose total benefit was equal to P. To find point B, Equity moved vertically upward to find a budget at the same cost as P but greater in benefit. However, there were none; B was the closest position. The flat portion at the top consisted of budgets that were more costly than B, but added no benefit, so Equity disallowed them. Any budget in the darker shaded area would thus be better than the present allocation of resources: more benefit for less cost. Of particular note were the budgets shown as points from C to B, inclusive. All of these are better value for money than P.

Figure 7.23 Children and families: Preferences and costs



In this workshop many of the savings options were associated with better ways of delivering the service, so they were given positive benefit scores even though they were less costly. Coupled with this, most services had offered only a few new development options, so Equity could not find any further options to buy. This is why the budgets shown on the upper surface are all characterised by negative costs. Given that the task was to find 5 per cent of savings, after years of previous cuts in local government, the people who developed the options appear to have found it difficult to contemplate new service developments. This is a typical finding when budgets have shrunk over the years.

However, as the task facing the group was to explore the impact of reducing the current budget by £1.3 million, the group moved on to consider this problem without exploring the 'B' and 'C' packages.

In the first instance the model was used by the group to list all savings options that either generated benefit or which had no dis-benefit, and only mandatory growth. This enabled the group to identify the first £840K of the savings target. At this stage a comment was made that *'while the savings were all "do-able", many of the options required new arrangements with existing and new partners and a considerable amount of management time to achieve'*.

The next stage was to use the model to identify which savings options would be recommended in order to meet the savings target in full.

Finally the group used the model to consider the service development (or Growth) options. Like the savings options these had also been prioritised into their 'Value for Money' sequence. The question the group was now faced with was *'...are you prepared to make even further savings in order to fund the few service developments on offer?'* The model was used to indicate those service development options that gained more Value for Money than savings options lost, i.e. where the gain was greater than the pain. It was interesting to note that all but one of the service developments was considered better Value for Money than the savings options offered. In other words, the group was able to identify how their service developments could be funded even within their tight financial constraints.

7.5.8 SENSITIVITY ANALYSIS

The alternative set of criteria weights favoured by the minority of the participants was fed into the model. This produced a slightly different prioritised list of options. However at the target level of savings the recommendations were exactly the same. It became apparent to the group that the weights under discussion were not the issue and that they could all agree to the model's recommendation.

7.5.9 OUTCOMES

At the end of the workshop the participants made the following comments:

"Process has allowed us to begin challenging the way the budget was built"

"Process only as good as options put in"

"Process exposed new options"

"We may underestimate the cumulative impact of a series of cuts"

"Rigorous, focussed"

"Process has forced us to face issues"

"Pre-meeting to review options could have reduced some of the debate"

"£1.3 million cut very much in line with what we would have expected"

"Useful in forcing us to examine our budgets in more detail than we otherwise would have"

"Found ways of making better use of my own money"

"Gives us a nice structure to work around"

The recommendations from the workshop were subsequently taken forward to the politicians who make the final decision. Their initial response was on the whole positive. The corporate directors for the whole authority are now interested in widening the use of the process.

7.6 Lessons from the MCDA case studies

Every new application of MCDA provides opportunities for learning how to improve the process next time. Some of the lessons are sector specific, so given the modest application of MCDA in the UK's public sector, much remains to be learned. Many of the paper and book references in the bibliography report UK experience, so those would be a good starting point for the reader who intends to apply MCDA. Satisfactory applications of MCDA require as much art to be exercised as science, but there is no point in re-inventing the wheel.

That said, the MCDA case studies reported here illustrate several general issues that will help in implementing MCDA.

1. MCDA carried out very early in the life of a new project can usefully guide the search for further information. The first attempt at modelling will highlight many inadequacies, in identifying and defining options and criteria, in the provision of data, in the inability to agree scores, and in judgements of trade-offs. At this point, the newcomer to MCDA may become discouraged, but, take heart, this is a good sign, for it identifies areas where further work is required. Thus, the MCDA modelling process provides an agenda for systematically tackling difficult issues, and shows which issues are important because their resolution could affect the overall result. The first one-day workshops for both Nirex and the NAO revealed areas of missing data, which were provisionally filled by participants' rough judgements. The evolving model helped to focus efforts between successive workshops on collecting relevant data, so that by the last workshop the models were complete and resources had not been wasted collecting data that would have been irrelevant.
2. Context matters. The Nirex case showed that the MCDA would have been carried out differently if the purpose had been final site selection rather than the creation of a short list of sites for further investigation to recommend to the Nirex Board. What is the purpose of the MCDA? This is the very first question to ask, for the answer is not always directly to support a decision. The MCDA usually contributes in some way to the decision making process, but the other possibilities listed in section 6.2.3 may require different approaches to the MCDA process.
3. MCDA is a socio-technical process. The technical apparatus provided by the theory of decisions with multiple objectives forms only a part of the MCDA process. Designing the social process within which the technical modelling will take place is crucial to success. The NAO cannot release any final report until the audited department agrees to the factual content of the report. Thus, it was important to obtain the co-operation of the DTI right from the start in the export study. The presence of their staff in the workshops helped to construct a value tree that accurately reflected the DTI's objectives for export services, ensured that DTI trade-offs were represented in the model, not those of the NAO, and so forth. For the Nirex case, key players were chosen to provide expertise about all the criteria. With the benefit of hindsight from the subsequent public enquiry,

the process might have been designed to elicit views from outside interest groups. Facilitated workshops were deemed appropriate for both cases: one-day sessions with several weeks between meetings to allow for data to be gathered. A single one-day session is rarely successful, for the approach is sufficiently novel that people require time to digest the work from each session, and they need subsequent sessions to revise and refine the model. The local authority anticipated a budget reduction of 5 per cent in real terms, so the focus of their efforts was on how to deliver more benefits at less cost.

4. By its nature, MCDA is an open, consultative process. Site selection in the Nirex case was the source of controversy at the public enquiry because the initial submission was incomplete. Only when full information was finally provided during the enquiry was it possible adequately to debate the MCDA, though various elements of controversy remained, for example because of the issue of incommensurability in value trees, the 'idiosyncrasy' of the NIREX team's weightings, and the fact that to this day the location of all the sites considered has not been revealed. MCDA reveals the value judgements that are a necessary part of any informed decision, so the social process must allow for the open expression of those views in the appropriate forum. Also, peer review may help to legitimise that expression.
5. Conduct the MCDA in an iterative fashion. There is no need to get every input to the model correct on the first go. Rely heavily on participants' judgements, even if they feel unsure. Subject vague inputs to sensitivity analyses, and find which inputs really matter to the overall results. Refine the value tree, scores and weights throughout the process. For both the NAO and Nirex cases, additional information was gathered between the facilitated workshops, and the model helped direct information gathering. Rely on participants' intuitions and gut feelings throughout the process. If results of the modelling don't feel right, explore the discrepancy. Sometimes intuition is wrong, while at other times the model requires revision. After several iterative cycles, the sense of unease should go away or diminish to the point where it is no longer an issue. The model can then be considered "requisite" in that it should be sufficient in form and content to resolve the issues at hand. In short, it is fit for purpose. That is the time to stop modelling. Don't over-complicate the MCDA model. Avoid analysis paralysis.
6. Leave time to explore the model fully. MCDA is a 'divide and conquer' strategy in the sense that a complex issue is subdivided into parts that are easier to deal with separately, and the theory is used to reassemble the parts. The resulting high-level view typically reveals new features that were not observed in the individual parts. Studying the properties of hydrogen and oxygen separately does not reveal the properties of water. Not until the benefit-cost displays were shown to the group in the NAO export study did it become clear which services were most and least cost efficient. Even the Jones family felt that they understood the toasters better after seeing the benefit-cost display in section 6.2.13. Creating different displays, changing scores to explore disagreements, doing sensitivity analyses on weights, all these help participants to gain a

better qualitative feel for the issues. That leads to increased confidence in taking a decision. The local authority case illustrates an important feature of MCDA modelling: its insensitivity to many differences of opinion. The criteria weights favoured by a minority of participants gave a slightly different prioritisation, but exactly the same recommendations for the target level of savings. Thus participants could agree about what to do without agreeing on all the input data. This is an important feature of using numbers to express judgements.

7. People make decisions, not models. MCDA models can assist people in making decisions, but the assistance can take many different forms: providing structure to debates, ensuring quality conversations, documenting the process of analysing the decision, separating matters of fact from matters of judgement, making value judgements explicit, bringing judgements about trade-offs between conflicting objectives to the attention of decision makers, creating shared understanding about the issues, generating a sense of common purpose, and, often, gaining agreement about the way forward. MCDA can do any or all of these, but it doesn't give 'the' answer. In particular, people looking for 'objective' methods for decision making are typically sceptical of MCDA, and may distrust the subjective judgements that are necessary inputs to the modelling. They do not recognise that there is no theory of objective decision making, that decision making is necessarily a human function.⁵⁹ The methods covered in this manual draw on decades of psychological research showing how it is possible to elicit from people judgements that are precise, reliable and accurate.⁶⁰
8. MCDA output should be framed to be in a form to best serve the purpose of the analysis. The report of the NAO export study showed the value tree and a plot of the services in a benefit versus cost space. Most of the results were described in words. In the Nirex study, the final report of the MCDA sessions reported the overall results and recommendations in just three pages with no graphs. Sensitivity analyses enabled the results to be couched as conditional conclusions: if you believe this, this and this, then Sellafield is best, but if you believe that, that and that, then Dounreay is best. This approach highlights the key value judgements, providing realistic freedom of choice, within bounds, for the decision maker. Only the recommendations from the local authority MCDA were taken forward to the politicians who were the final decision makers.

⁵⁹ Those who aspire to 'objective' justifications for public policy decision making must contend with the profound theoretical difficulties of interpersonal comparison of preference intensity and the infamous Arrow Impossibility Theorem for ordinal preference orderings. The notion of a single discrete 'objective' social preference ordering is theoretically weak and unlikely to be achieved in practice in a pluralistic society.

⁶⁰ See, for example, Wright, G., & Bolger, F. (Eds.). (1992) *Expertise and Decision Support*, Plenum Press, New York.

Appendix 1 Continuous MCA models (infinitely variable alternatives)

This appendix clarifies the distinction between two main classes of multi-criteria decision problems. It indicates in outline how solutions to the class not discussed in the main text of the manual are obtained.

As indicated in section 4.4, most policy decisions concern choices between a finite number of options, the details of which have already been predetermined before they are subject to MCA. This does not exclude the possibility that, following MCA, the decision making group may use insights from the MCA, coupled with expertise from other sources, to re-define some options and run the MCA again. To do so would be quite normal. However, the basic MCA method itself does not re-define the options. It is concerned simply to assess the strengths and weaknesses of options as they stand.

There is, however, a second class of problems where the MCA methods themselves, sometimes using interactive computer methods to involve the analyst explicitly in the process also, directly seek to specify what the definition of the best option should be. The question is essentially one of identifying an optimal **design** for the option, guided by MCA methods. Almost always, the optimisation is subject to specific constraints, for example on cost or technical specification.

As a starting point, consider the basic linear programming (LP) problem from operational research (OR). It seeks to identify from all the (infinite) possible combinations of values of a set of decision variables, x_j , a set which maximises a given linear objective function while also obeying a set of constraints which restrict the combinations of x_j values that are admissible. The constraints are also all represented by linear functions and, in addition, the decision variables are required to take only non-negative values.

Maximise

$$\sum_{j=1}^n a_j x_j$$

Subject to

$$\begin{aligned} \sum_{j=1}^n a_{ij} x_j &\leq b_i \quad (i = 1, \dots, m) \\ x_j &\geq 0 \quad (j = 1, \dots, n) \end{aligned}$$

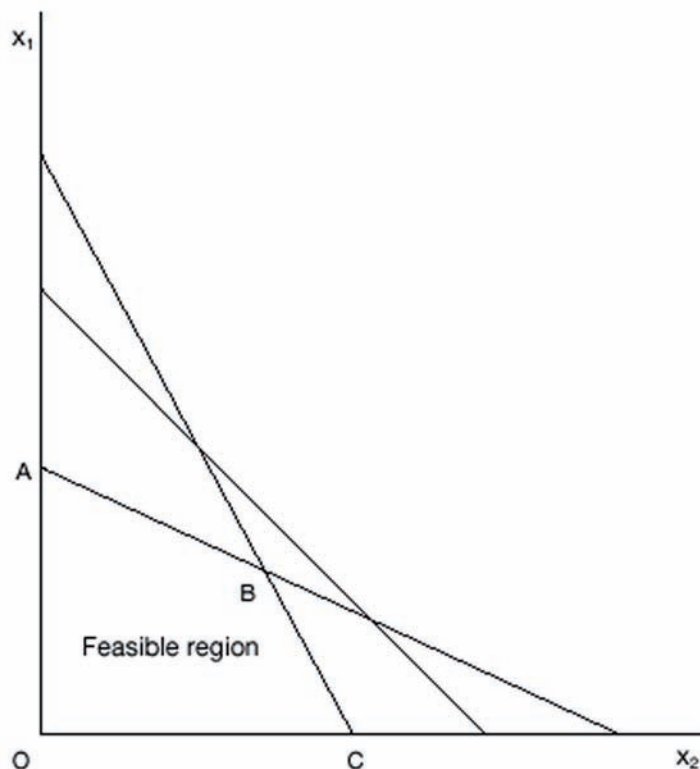
The x_j are decision variables – variables over which the decision maker has control. The a_j are numerical parameters whose relative values reflect the relative contributions of changes in each of the decision variables to achieving the overall aims.

$\sum a_j x_j$ is the objective function – it expresses the decision maker's overall goal (objective) as a function of the decision variables.

$\sum a_{ij} x_j \leq b_i$ are functional constraints – they express how the values of the x_j are limited by the operating environment in which the decision maker finds him/herself

$x_j \geq 0$ are non-negativity constraints, requiring that the x_j do not take negative values (the possibility of negative x_j can be accommodated indirectly).

A classic example is the product mix problem. Here the x_j represent quantities of different goods a firm may choose to manufacture, the a_j in the objective function are unit profit rates on the goods and each of the constraints reflects the maximum amount of a critical input that is available (e.g., skilled labour, machine time). Each good requires the use of a_{ij} units of scarce input i per unit of output. The overall LP problem is to design the optimal product mix (set of x_j values) to maximise total realised profit (the objective function). Examples of this type feature in nearly all introductory OR texts.⁶¹ If the problem is restricted to just two decision variables, it can be represented graphically:



⁶¹ For example, Anderson, D.R., Sweeney, D.J. and Williams, T.A. (1997) *An Introduction to Management Science*, (8th edition) West Publishing Company, St. Paul, Minnesota.

The aim is to find the point within the feasible region delimited by the combination of all the constraints that allows the objective function to take its maximum value. It is a key result of LP theory that there will always be an optimal solution at one of the corner points of the feasible region.

In LP, a little counter-intuitively, the objective function is in some ways the **least** important consideration. The constraints **must** be obeyed. With the objective function, the aim is simply to do the best possible, provided the constraints are not broken. Sometimes it is not obvious whether an aspect of a particular problem should be represented as a constraint, or should be relegated to being the objective function.

In real life application it is not uncommon to be facing several objectives at once with no obvious way of deciding that one should be **the** objective function and the rest be represented as constraints. For example, a firm may be interested simultaneously in maximising profit, minimising environmental damage, maximising market share, minimising capital equipment cost, and so on.

Problems of this type, where the decision variables are infinitely variable, subject to constraints and where there are multiple objectives, are often called multiple objective decision making problems (MODM). Note that, in general, neither the constraints nor the objective functions need be linear, although this is often assumed. Where the problem is linear, its representation is just like that of the LP problem above, except that there are now several objective functions, not just one.

Solution methods for MODM problems are often strongly problem-specific. However, a common feature of the search process for a solution is a concern to identify **efficient** solutions. This reflects thinking similar to the ideas behind dominance (section 5.5.2.1). The aim, broadly, is to restrict consideration to a (relatively small) set of solutions to the MODM problem that are explicitly not dominated by others. MODM techniques therefore start by implementing a search for efficient solutions. Once these have been identified, they then move into a second phase in which different solutions are explicitly compared using the range of objective functions that are defined for the problem in question. Often this will involve explicit interaction with the user, who, directly or indirectly, will ultimately have to input views on the trade-offs or other measures of relative importance that he/she holds as between the competing objectives. Often the solutions found in this way cannot be *proved* to be optimal; the best that can be hoped for is that they are relatively good.

Details of MODM procedures are beyond the scope of this manual and the solution of MODM problems is likely to lie outside the field of interest of most public sector decision makers. For further information, Steuer⁶² is an excellent starting point. The *Journal of Multi Criteria Decision Analysis* is a good source for more recent developments.

⁶² Steuer, R.E. (1986) *Multiple Criteria Optimization: Theory, Computation and Application*, John Wiley, New York.

Appendix 2 Non-compensatory methods

This appendix outlines procedures that exist to try to establish preferences between options when:

- each option is evaluated against a common set of criteria set out in a performance matrix
- the decision maker is not willing to allow compensation, ie, for strong performance on one criterion to compensate for weak performance on some other criterion.

Commitment to non-compensatory evaluation severely restricts the extent to which, in practice, overall preferences between options can be established.

2.1 Dominance

As indicated in section 5.5.2.1, the only directly applicable way of undertaking any holistic comparison of options based solely on non-compensatory analysis of the performance matrix is to search for dominance. Typically in practice, few cases of dominance will be present, even at the level of an option being dominated by one or two others. The likelihood of an option dominating or being dominated by **all** others is very small. Being dominated would (all else being equal) constitute a reason for not considering the option any further. Similarly, if an alternative dominates all others, it is the best available.

Dominance assessments (being based ultimately on a series of yes/no calculations) can be quite sensitive to errors in the data in the performance matrix. They are also potentially changed by the addition or removal of even a single option to or from the set under consideration.

2.2 Conjunctive and disjunctive selection procedures

If the decision maker is prepared to allow one or more supplementary judgements to complement the information in the performance matrix, then it is possible to go a little further with non-compensatory decision making about options.

One possibility is to allow the introduction of thresholds of performance for one or more criteria. The extra information introduced consists of the thresholds themselves and, implicitly, a judgement that the criteria concerned

justify being prioritised in this way over others for which no thresholds may be given. What is effected is a form of external benchmarking.

The conjunctive model eliminates options that fail to reach externally set levels of performance on each of one or more named criteria. For example, candidates to be short-listed for a job interview might be expected to have reached certain minimum levels of achievement on each of a number of the criteria used in the selection process.

The disjunctive model allows an option to pass if the option concerned meets a minimum threshold level of performance on at least one of a set of named criteria. Continuing the selection example, this would be equivalent to allowing candidates through to the short-list provided they exhibited potential (i.e., exceeded the selected threshold) in terms of any one of the key criteria for which thresholds have been set.

It is perfectly permissible to use both a conjunctive and a disjunctive filter in any one decision making process, with different threshold levels, perhaps applied to different sets of criteria.

If the external benchmarks are set independently and represent accepted reasons (e.g., a safety threshold) whereby an option that fails to 'pass' would not normally ever be considered for implementation, then this type of screening process is valid, although it is arguable that it scarcely needs formal MCA to make the judgement. If, on the other hand, the performance thresholds are being set in the light of knowledge of what all the options are scoring in this particular analysis, then the process may induce unforeseen bias. By selecting particular criteria to be used for benchmarking (or by identifying benchmark levels themselves) an implicit judgement is being made about the relative importance of criteria, the basis for which may not have been fully thought through.

General practical advice would be that disjunctive and conjunctive procedures may help in providing some structure and an audit trail in developing a long short-list from an initial list of candidate options. However, for later stages in the decision making process, where candidate options are harder to separate in terms of overall performance, the MCDA model is likely to prove a more reliable guide.

2.3 Lexicographic ordering

Another approach to non-compensatory choice requires the decision-maker not to supply external benchmarks, but instead to provide supplementary information about the ranking of criteria in terms of perceived importance. It then considers each criterion in turn and works as a sequential elimination model.

Specifically, in lexicographic elimination, all options are first compared in terms of the criterion deemed most important. If there is a unique best

performing option in terms of this criterion, then that option is selected as the most preferred. If there is a tie, then the selection process moves on to the second ranked criterion and, just for those options which tied for first previously, seeks the remaining option which scores best on the second criterion. Again, if this leads to a unique selection, then this option is designated the preferred one. If not, the process is repeated for options tying in terms of both the first and second ranked criteria, using the third criterion and so on until a unique option is identified or all criteria have been considered.

2.4 Elimination by aspects

This model combines elements of both lexicographic ordering and the conjunctive/disjunctive models. Options are compared against a threshold. They are examined criterion by criterion and, for each criterion, options which do not pass the threshold are eliminated. However, criteria are examined not in order of importance, but in perceived order of likelihood of maximising the number of options that fail to pass. This process is continued until only one option remains.

Neither lexicographic elimination nor elimination by aspects have contributed much to the practice of public sector decision making.

Appendix 3 Multi-attribute utility models

Research into decision making is divided into three principal streams:

- Descriptive: which examines how individuals, groups and organisations actually undertake decision making in practice
- Normative: which tries to establish how rational individuals **should** choose between competing options, and
- Prescriptive: which, recognising some of the weaknesses in intuitive, unaided human decision making that descriptive decision research has identified, seeks procedures to bring decision making in practice closer to normative ideals.

Most of this manual concentrates on prescriptive⁶³ approaches to decision making. In this section, however, the initial focus is normative.

Although there are some clear links back to earlier work, mostly on the analysis of economic behaviour and decision making, the starting point for MCA in terms of a normative theory of how individuals **should** rationally choose between competing options is generally seen as the work of von Neumann and Morgenstern,⁶⁴ followed by Savage.⁶⁵

These authors aimed to derive a theory of how rational individuals ought to choose between options. They first established a set of fundamental axioms of rational choice (e.g., one axiom is that more of a desirable good should be preferred to less of it). Then, by using mathematical reasoning, they showed that the only way an individual could behave consistently with the full set of axioms is by choosing the option which possessed the maximum subjective expected utility (SEU) value.

It is assumed that there is a range of separate options to choose between, that one and only one of these options must be chosen now, and that, because of uncertainty about exactly what the future will be, different options will have potentially different values (utilities) to the decision maker, depending on what kind of future (termed state of nature, or state of the world) eventually transpires. For example, the utility of investing in a new reservoir may depend on future rates of climate change.

⁶³ Note that **prescriptive** may be interpreted in either a strong or a weaker sense. Prescription in the sense of an order or instruction to the decision-maker is **not** what is implied here. The procedures in this manual should be interpreted in the weaker sense as support or advice for the decision-maker.

⁶⁴ Von Neumann, J, and Morgenstern, O. (1947) *Theory of Games and Economic Behaviour*, second edition, Princeton University Press, Princeton.

⁶⁵ Savage, L.J. (1954) *The Foundations of Statistics*, Wiley, New York, NY.

The SEU of any option is obtained by:

- (i) identifying all future states of the world that could reasonably be viewed as relevant to this decision
- (ii) calculating the utility (degree of attractiveness), u_{ij} , the decision maker associates with the outcome that follows from the combination of choosing option i (say) and it later turning out that future state of the world j actually occurs
- (iii) creating the probability weighted average of all the outcome utilities, where the probabilities are the individual's subjective estimates of the probability of each of the outcomes actually occurring.

$$U_i = p_1u_{i1} + p_2u_{i2} + \dots + p_nu_{in} = \sum_{j=1}^n p_ju_{ij}$$

Where:

U_i is the overall utility (preference score) of option i ;

u_{ij} is the utility of option i if, having chosen option i , it subsequently transpires that state of the world j occurs;

p_j is the decision maker's best judgement of the probability that future state of the world j will occur.

This basic model of rational choice under uncertainty, although not without its critics, is the single most broadly accepted normative model of rational choice. While it deals explicitly with uncertainty, it is also implicitly a multi-criteria model since each of the individual u_{ij} are in principle based on a multi-criteria assessment.

The SEU model, however, does not provide direct prescriptive decision support for the multi-criteria problems addressed in this manual. This is because it does not indicate how the u_{ij} should be evaluated. Progress in this respect did not occur until the publication of Keeney and Raiffa which presented a set of procedures for operationalising the calculation of (what they termed) multiattribute utilities.⁶⁶

It is beyond the scope of this manual to go into full detail on procedures for calculating u_{ij} estimates, which can be relatively complex. For this, see, eg Keeney and Raiffa⁶⁷ or Goodwin and Wright⁶⁸ for an introductory treatment. The extra complexity stems from the fact that the multiattribute utility model is simultaneously seeking to take account both of uncertainty and evaluating in terms of several criteria.

In calculating utility scores a critical issue is mutual independence of preferences. If mutual independence can be established, then the calculation

⁶⁶ Their use of the term attribute is equivalent to what this manual terms criterion.

⁶⁷ Keeney, R. L. and Raiffa, H. (1976) *Decisions with Multiple Objectives: Preferences and Value Trade-Offs*, Wiley, New York.

⁶⁸ Goodwin, P., and Wright, G. (1988) *Decision Analysis for Management Judgement* John Wiley, Chichester.

of the individual u_{ij} can be done using methods not a great deal more demanding than the MCDA technique set out in Chapter 5. This would not be an unreasonable task, even for a non-specialist. However, if this is not the case, the mathematical structure of the equation to calculate u_{ij} will be more difficult.

One response can be to try to restructure the MCA, using re-defined criteria that do exhibit mutual independence of preferences. The alternative is to use non-linear equations to express (the multi-criteria) u_{ij} in terms of utility scores for performance levels on each of the individual performance criteria.

For example, in an early MCA of a blood bank inventory control problem Keeney concluded that the overall utility associated with different operating policies depended critically on just two underlying performance dimensions, x which relates to the possibility of not being able to supply a particular patient's blood type when it is needed, and y , which relates to the amount of blood ordered for inventory that exceeds its "use-by" date before it is needed x (and so has to be wasted).⁶⁹

Following carefully structured and relatively detailed discussions with blood bank professionals, the authors concluded that independence of preferences between x and y did not exist in this case. Further, there was evidence that the utility function (equivalent in conditions of uncertainty to the value functions described in Chapters 4 and 5) linking utility to performance with respect to x and y individually was non-linear. The combination of this non-linearity and the lack of preferential independence led to an overall utility function of the form:

$$U(x,y) = 0.32(1 - e^{-0.11x}) + 0.57(1 - e^{-0.04y}) + 0.107(1 - e^{-0.11x})(1 - e^{-0.04y})$$

While this is not specially difficult in mathematical terms, it is an order of magnitude more demanding than the methods outlined in Chapter 5. Determining the functional form and estimating the parameters for a utility analysis of this type, even though it is an MCA with only two performance criteria, is perhaps beyond what most non-specialists would feel confident to do. Overall experience of seeking to operationalise the Keeney and Raiffa application of the underlying SEU model is that it most often undertaken by specialist consultants. This, in turn, has cost implications for the analysis that mean that multiattribute utility theory is usually applied only to problems where uncertainty levels are high and the overall financial and/or human consequences are high. Some of the most important and far-reaching applications recently have concerned decisions about the reprocessing or storage of nuclear waste.

Keeney and Raiffa's 1976 book is the key guide to multiattribute utility applications. Other references that cite examples and indicate the range of potential application include Corner and Kirkwood⁷⁰ and Keeney.⁷¹

⁶⁹ Keeney, R.L. (1972) 'An illustrated procedure for assessing multi-attributed utility functions' *Sloan Management Review Fall*, 37–50.

⁷⁰ Corner, J.L., and Kirkwood, C.W. (1991) 'Decision analysis applications in the operations research literature, 1970–1988' *Operations Research*, 39, pp.206–219.

⁷¹ Keeney, R.L. (1992) *Value-Focused Thinking: a Path to Creative Decision-Making* Harvard University Press, Harvard, MA.

Appendix 4 Linear additive models

4.1 Introduction

The linear additive multi-criteria model set out in section 4.7 has a straightforward intuitive appeal and transparency that ensures it a central role in any discussion of MCA. In part for this reason, it provides the basis for the MCA and MCDA procedures discussed in this manual.

However, as with many tools, it can be misused. Its very ease of use can encourage this. It is important to follow the steps set out in Chapters 5 and 6 with care. In particular, this is critical with regard to the scaling of options' performances on criteria, the weighting of criteria, and the relationship between weight determination and the scales on which performance on each criterion is measured (see section 6.2.10). Failure to follow the proper logic of the model can lead to an MCDA that appears clear and well-founded, but which is, in fact, misleading and not a true reflection of the decision making group's understanding of the problem.

4.2 Background

Perhaps the earliest example of a linear additive MCA is the **Prudential Calculus** proposed by Benjamin Franklin to Joseph Priestley (see Table A4.1).

The weighting and scoring procedure used here falls a little short of what may be required for public sector decision making, but the method is not without its merits. Like many MCA procedures, it has the effect of encouraging the decision maker to think carefully about identifying key criteria.

4.3 More recent developments

More recently than Franklin, a number of researchers have developed the linear multiattribute model and ways of applying it that are helpful in various circumstances. Two important perspectives in this respect derive, respectively, from Keeney and Raiffa⁷² and Edwards.⁷³

⁷² Keeney, R.L. and Raiffa, H. (1976) *Decisions with Multiple Objectives: Preferences and Value Trade-offs*, Wiley, New York, reprinted, Cambridge University Press, 1993.

⁷³ Edwards, W. (1971) 'Social utilities', *Engineering Economist*, Summer Symposium Series 6, pp.119–29

The importance of the Keeney and Raiffa work has already been alluded to in section 4.6 and Appendix 3. By showing that, in many circumstances, a linear additive model:

$$S_i = \sum_{j=1}^n w_j s_{ij} = w_1 s_{i1} + w_2 s_{i2} + \dots + w_n s_{in}$$

can be a robust and straightforward approximation to the von Neumann and Morgenstern, SEU model, with its rigorous axiomatic foundation, Keeney and Raiffa provide a powerful legitimation for its practical application. This is especially so in circumstances where accurate prediction of alternatives performances may be difficult, where the aim is to screen options to develop a shortlist and/or where broader political or other considerations will fine-tune the ultimate choice. Using a simple, transparent model in such circumstances, probably coupled with some sensitivity testing, often provides a very good return for the analytical effort consumed.

Independently, Edwards reached a similar conclusion. Building on work in which psychologically-oriented decision researchers had been trying to build models to reflect how expert decision makers make decisions, he came to develop *SMART* (Simple Multi-Attribute Rating Technique, which is not to be confused with the term referred to in Section 2.3). *SMART* was originally a seven-step procedure, with steps very similar to those set out in Figure 5.1, which allows a decision making group straightforwardly to set up a MCDA model.

Table A4.1 Benjamin Franklin's letter

Dear Sir,

London Sept. 19. 1772

In the Affair of so much Importance to you, wherein you ask my Advice, I cannot for want of sufficient Premises, advise you what to determine, but if you please I will tell you how. When these difficult Cases occur, they are difficult chiefly because while we have them under Consideration all the Reasons pro and con are not present to the Mind at the same time; but sometimes one Set present themselves, and at other times another, the first being out of sight. hence the various Purposes of Inclinations that alternately prevail, and the Uncertainty that perplexes us. To get over this, my Way is, to divide half a Sheet of Paper by a Line into two Columns, writing over the one pro, and over the other Con. Then during three or four Days Consideration I put down under the different Heads short Hints of the different Motives that at different Times occur to me for or against the Measure. When I have thus got them all together in one View, I endeavour to estimate their respective Weights; and where I find two, one on each side, that seem equal, I strike them both out: If I find a Reason pro equal to some two Reasons con, I strike out the three. If I judge some two Reasons con equal to some three Reasons pro, I strike out the five; and thus proceeding I find at length where the Balance lies; and if after a Day or two of farther Consideration nothing new that is of Importance occurs on either side, I come to a Determination accordingly. And tho' the Weight of Reasons cannot be taken with the Precision of Algebraic Quantities, yet when each is thus considered separately and comparatively, and the whole lies before me, I think I can judge better, and am less likely to make a rash Step; and in fact I have found great Advantage from this kind of Equation, in what may be called Moral or Prudential Algebra. Wishing sincerely that you may determine for the best, I am ever, my dear Friend, Yours most affectionately.

B Franklin

Dr Priestly

Subsequent to its initial formulation and early applications, *SMART* has undergone one important change and one potentially useful extension. The change was to introduce the swing weights procedure for weight assessment (see section 6.2.10) to replace the one originally recommended. Thus *SMART* became *SMARTS* (*SMART* with Swings).

Secondly, Edwards and Barron⁷⁴ have also developed a way of implementing *SMARTS* that is less demanding in its information input requirements from the decision maker. This procedure is named *SMARTER* (*SMART* Exploiting Ranks).

Although the MCDA model is not difficult to apply, for some time researchers have sought ways of using linear additive choice models that require less precise data inputs than the basic model. Of course, with less precise information input, the recommendations output by the model will be less precise too and less likely unambiguously to identify a 'best' option. However, it may be that this is a price worth paying, if the input demands on the decision maker are less, or if the model can be constructed and applied more quickly.

The most common way to seek to accommodate less precise inputs is to allow specific weights or scores to be replaced by statements that put bounds on the values the inputs can take, and/or express restrictions on the relative magnitude of some of the input values. These restrictions are most often in the form of linear inequalities (see below).

Early thinking along these lines was developed by Fishburn⁷⁵ in the context of single criterion decision making under uncertainty (circumstances mathematically analogous to the linear additive MCA model). This work was subsequently developed in a number of ways by several authors. Examples include: Kofler et al.,⁷⁶ Weber⁷⁷ Pearman^{78,79}, Park et al.⁸⁰ and Athanassopoulos and Podinovski.⁸¹

In the linear additive MCA model, there are two inputs, weights and scores. In principle, decision makers may well be uncertain about the accuracy of either w_j or s_{ij} , or both. In many cases but not all, it is the weights where confidence about accuracy is lower. This is because many of the criteria will be assessed on objective scales of measurement, or against a shared background of evidence and experience that will tend to focus decision makers towards broadly similar assessments. For example, in assessing

⁷⁴ Edwards, W. and Barron, F.H. (1994) 'SMARTS and SMARTER: improved simple methods for multiattribute utility measurement', *Organizational Behavior and Human Decision Processes*, 60, 1994, pp.306–25.

⁷⁵ Fishburn, P.C., (1965) 'Analysis of decisions with incomplete knowledge of probabilities', *Operations Research*, 13, pp.217–37.

⁷⁶ Kofler, E., Kmietowicz, Z.W and Pearman, A.D. (1984) 'Decision making with linear partial information (LPI)', *Journal of Operational Research Society*, 35, pp 1079–90.

⁷⁷ Weber, M. (1987) 'Decision making with incomplete information', *European Journal of Operational Research*, pp.44–57.

⁷⁸ Pearman, A.D. (1991) 'The identification of potentially optimal strategies in LPI decision making', *Journal of Operational Research Society of India*, 28, pp.102–11.

⁷⁹ Pearman, A.D. (1993) 'Establishing dominance in multiattribute decision making using an ordered metric method', *Journal of the Operational Research Society*, 44, pp.461–6.

⁸⁰ Park, K.S., Kim, S.H. and Yoon, W.C. (1996) 'An extended model for establishing dominance in multiattribute decisionmaking', *Journal of the Operational Research Society*, 47, pp.1415–20.

⁸¹ Athanassopoulos, A.D. and Podinovski, V.V. (1997) 'Dominance and potential optimality in multiple criteria decision analysis with imprecise information', *Journal of the Operational Research Society*, 48, pp.142–50.

a school's ambience as experienced by new pupils or teachers, most experienced inspectors would probably look for similar indicators and reach broadly similar conclusions in assessing a particular school. Where there may well be disagreement, however, is in the relative importance of this performance indicator in relation to others contributing to the overall standing of an institution.

For this reason, typically the scores are treated as known and formal analysis of imprecision is concentrated on w . However, in some areas of public policy debate, especially where there are diverse groups of stakeholders, there may well be substantial disagreement about scores. It is possible to apply similar methods to those described below with the weights fixed and scores allowed to vary, for example between upper and lower bounds, or to allow both weights and scores to vary simultaneously. These two possibilities are discussed at the end of this appendix.

Uncertainty about the precise value of the weights may be reflected in two main ways. First, some or all of w may be ranked, eg:

$$w_1 \geq w_2$$

$$w_5 \geq w_2 + w_4.$$

Alternatively, limits may be placed on the proportion of the overall weight in the decision that particular criteria may have:

$$w_3 \leq 0.25$$

$$0.2 \leq w_4 \leq 0.5$$

Note that in each case what is being compared is the weight to be associated with a swing from the lowest to the highest considered levels for each of the criteria concerned – in other words, the same thought process as underlies swing weighting.

Bearing in mind that the sum of the weights must always be one, linear inequality restrictions such as the above combine to restrict the values that the weights can take, while not going so far as to fix precise numerical values for them. Thus the full extent to which the decision maker is prepared to make commitments about the relative importance of the different criteria is captured, without requiring an input of more precise estimates than he/she is happy to give.

The linear inequalities define what is in effect a feasible region (see Appendix 1) which specifies every possible combination of weight values that could be consistent with the decision makers' stated views. There are then two ways this information can be used.

One approach is to concentrate on the maximum and minimum values the weighted average S_j scores can take for the various options. Linear programming theory establishes that these extreme values will occur at vertex

(corner) points of the (n -dimensional) feasible region. Procedures exist to identify the full set of w_j values at these vertices. Depending on the nature of the inequality constraints, the values may be determined by direct application of a mathematical formula or by using a specialised computer package (e.g., the vertex enumeration algorithm included in the *Mathematica* package).

Once the vertex points are identified, it is a simple matter to:

- find the maximum weighted value each option can take
- find the minimum weighted value each option can take
- look for dominance – here, the possibility that the minimum weighted difference between two options is greater than zero.

The extent to which this process will finalise the recommendation of a single option will depend on how restrictive are the set of inequality restrictions on w_j and on the relative performance of the options. Often it will not be sufficient to determine a unique recommendation. Dominance in weighted score differences may eliminate some options. Thereafter a more judgmental process may be needed. For example, options might be separately ranked in terms of their maximum and minimum weighted values. Then, a shortlist could be created from those options that offer both relatively high maximum weighted values and simultaneously avoid low minimum weighted values.

A second approach is to seek to identify a single set of weights that is representative of all the possible weight combinations that are admissible, consistent with the established linear inequality constraints on the weights. Barron and Barrett and others have argued that this point should be the centroid of the feasible region of admissible sets of weight values.

Thus, taking this approach, having identified the centroid, all that is necessary is to evaluate each option at the centroid point and to rank options in terms of the weighted average score they record at that point.

Edwards and Barron give a straightforward formula for determining the centroid point for the case where all criteria are ranked simply in the form $w_1 \geq w_2 \geq w_3 \geq \dots \geq w_n$ (which they term the Rank Order Centroid, ROC). Specifically, if there are n criteria, the weight for the j th criterion should be:

$$w_j = (1/n) \sum_{i=1}^n (1/i)$$

Deriving the centroid is more complex where the inequalities are less straightforward, but various procedures are available that permit the calculations to be made – see Barron and Barrett,⁸² Kmietowicz and Pearman.⁸³

⁸² Barron, F.J. and Barret, B.E (1996) 'Decision quality using ranked attribute weights', *Management Science*, 42, pp.1515–23.

⁸³ Kmietowicz, Z.W. and Pearman, A.D. (1984) 'Decision theory, linear partial information and statistical dominance', *Omega*, 12, pp.301–99.

Edwards and Barron suggest, on the basis of simulation studies, that ranking options using the ROC procedure successfully locates the best available option between 75 and 85 per cent of occasions. More importantly, when the best option is **not** located, the average loss of S score is only 2 per cent. Given likely inaccuracies in the data, this suggests that the ROC procedure can be a very effective short-listing device.

The above assumes that decision makers' weight estimates are imprecise, but that estimates of options' scores on each criterion are accurate. If the reverse is the case, then the first of the exploration and short-listing process described above (based on minimum and maximum weighted S scores and dominance in weighted scores) may be worth pursuing. In this case, however, it will generally be necessary to run a simple linear programming computer package to identify the required information about options' performance at the vertex points.

As an alternative, insights about variability in the overall performance levels of different options can also be obtained by applying simple risk analysis methods, where imprecision in the true values of individual performance scores, s_{ij} , can be represented by probability distributions. Computer packages such as Risk make this straightforward to undertake. This process, however, does not provide any direct means to rank options, which would normally be done subsequently by inspection and expert judgement.

Because of the extra restriction that weights must be non-negative values adding to one, applying risk analysis methods to explore the consequences of imprecision in the weights is not a simple thing to do.

Examining the consequences of **simultaneous** imprecision in both weights and scores is more complex. Moreover, the doubly uncertain set of inputs is likely in many cases to lead to output (S_i) performance ranges for options that are wide and overlapping, and, consequently, uninformative. In general, pursuing this line of enquiry is unlikely to be worthwhile.

Appendix 5 The analytical hierarchy process

5.1 The basic ahp procedure

At the core of the Analytic Hierarchy Process (AHP) lies a method for converting subjective assessments of relative importance to a set of overall scores or weights. The method was originally devised by Saaty.⁸⁴ It has proved to be one of the more widely applied MCA methods, see, for example, Zahedi,⁸⁵Golden et al.⁸⁶and Shim⁸⁷for summaries of applications. However, at the same time, it has attracted substantial criticism from a number of MCA specialists. There have also been attempts to derive similar methods that retain the strengths of AHP while avoiding some of the criticisms.

The fundamental input to the AHP is the decision maker’s answers to a series of questions of the general form, ‘How important is criterion *A* relative to criterion *B*?’. These are termed pairwise comparisons. Questions of this type may be used to establish, within AHP, both weights for criteria and performance scores for options on the different criteria.

Consider firstly the derivation of weights. It is assumed that a set of criteria has already been established, as discussed in Chapters 5 and 6. For each pair of criteria, the decision-maker is then required to respond to a pairwise comparison question asking the relative importance of the two. Responses are gathered in verbal form and subsequently codified on a nine-point intensity scale, as follows:

How important is A relative to B?	Preference index assigned
Equally important	1
Moderately more important	3
Strongly more important	5
Very strongly more important	7
Overwhelmingly more important	9

2, 4, 6 and 8 are intermediate values that can be used to represent shades of judgement between the five basic assessments.

If the judgement is that *B* is more important than *A*, then the reciprocal of the relevant index value is assigned. For example, if *B* is felt to be very

⁸⁴ Saaty, T. (1980) *The Analytical Hierarchy Process*, John Wiley, New York.
⁸⁵ Zahedi, F. (1986) ‘The analytic hierarchy process: a survey of the method and its applications’, *Interfaces*, 16, pp.96–108.
⁸⁶ Golden, B., Wasil, E. and Harker, P. (eds.) (1989) *The Analytic Hierarchy Process: Applications and Studies*, Springer Verlag, New York.
⁸⁷ Shim, J.P.(1989) ‘Bibliographical research on the analytic hierarchy process (AHP)’ *Socio-Economic Planning Sciences*, 23, pp.161–7.

strongly more important as a criterion for the decision than A , then the value $1/7$ would be assigned to A relative to B .

Because the decision maker is assumed to be consistent in making judgements about any one pair of criteria and since all criteria will always rank equally when compared to themselves, it is only ever necessary to make $1/2 n(n - 1)$ comparisons to establish the full set of pairwise judgements for n criteria. Thus a typical matrix for establishing the relative importance of three criteria might look like:

$$\begin{bmatrix} 1 & 5 & 9 \\ 1/5 & 1 & 3 \\ 1/9 & 1/3 & 1 \end{bmatrix}$$

The next step is to estimate the set of weights (three in the above example) that are most consistent with the relativities expressed in the matrix. Note that while there is complete consistency in the (reciprocal) judgements made about any one pair, consistency of judgements **between** pairs is not guaranteed. Thus the task is to search for the three w_j that will provide the best fit to the 'observations' recorded in the pairwise comparison matrix. This may be done in a number of ways.

Saaty's basic method to identify the value of the weights depends on relatively advanced ideas in matrix algebra and calculates the weights as the elements in the eigenvector associated with the maximum eigenvalue of the matrix. For the above set of pairwise comparisons, the resulting weights are:

$$w_1 = 0.751 \qquad w_2 = 0.178 \qquad w_3 = 0.070.$$

The calculations required are quite complex. In practice they would be undertaken by a special AHP computer package.

A more straightforward alternative, which also has some theoretical attractions (see below) is to:

- calculate the geometric mean of each row in the matrix
- total the geometric means, and
- normalise each of the geometric means by dividing by the total just computed.

In the example, this would give:

Geometric mean			Weight ⁸⁸
Criterion 1	$(1 \times 5 \times 9)^{1/3}$	3.5568	0.751
Criterion 2	$(1/5 \times 1 \times 3)^{1/3}$	0.8434	0.178
Criterion 3	$(1/9 \times 1/3 \times 1)^{1/3}$	0.3333	0.070
Sum		4.7335	(=1.00)

⁸⁸ Weights should sum to one. There is a small rounding error.

Taken to further decimal points of accuracy, the weights estimated by the two different methods are not identical, but it is common for them to be very close.

In computing weights, it is normal to cluster criteria in a value tree (see section 6.2.6). In AHP applications, this allows a series of small sets of pairwise comparisons to be undertaken within segments of the value tree and then between sections at a higher level in the hierarchy. In this way, the number of pairwise comparisons to be undertaken does not become too great.

In addition to calculating weights for the criteria in this way, full implementation of the AHP also uses pairwise comparison to establish relative performance scores for each of the options on each criterion. In this case, the series of pairwise questions to be answered asks about the relative importance of the performances of pairs of alternatives in terms of their contribution towards fulfilling each criterion. Responses use the same set of nine index assessments as before. If there are m options and n criteria, then n separate $m \times m$ matrices must be created and processed.

Although this may seem a daunting task, computer packages such as *Expert Choice*, and *HIPRE 3+* automate most of the computations. Generally, non-specialist users find the pairwise comparison data entry procedures of AHP and related procedures attractive and easy to undertake.

With weights and scores all computed using the pairwise comparison approach just described, options are then evaluated overall using the simple linear additive model used for MCDA. All options will record a weighted score, S_i , somewhere in the range zero to one. The largest is the preferred option, subject as always to sensitivity testing and other context-specific analysis of the ranking produced by the model.

5.2 Concerns about the AHP

The AHP provides the same benefits as do MCDA models in terms of focusing decision maker attention on developing a formal structure to capture all the important factors likely to differentiate a good choice of an option from a poor one. Pairwise comparisons are generally found to be readily accepted in practice as a means of establishing information about the relative importance of criteria and the relative performance of options. The fact that the pairwise comparison matrix provides some redundant information about relative values allows some cross-checking to be done. Arguably, the resulting weights or scores may be more stable and consistent than if they were based on a narrower set of judgements. AHP also fits comfortably with circumstances where judgements, rather than measurements of performance (say), are the predominant form of input information.

Nonetheless, despite these attractions, decision analysts have voiced a number of concerns about the AHP. French⁸⁹ provides a succinct critique; see also Goodwin and Wright.⁹⁰ The main doubts raised are:

- (a) The 1–9 scale has the potential to be internally inconsistent. A may be scored 3 in relation to B and B similarly scored 5 relative to C. But the 1–9 scale means that a consistent ranking of A relative to C (requiring a score of 15) is impossible.
- (b) The link between the points on the 1–9 scale and the corresponding verbal descriptions does not have a theoretical foundation.
- (c) Weights are elicited for criteria before measurement scales for criteria have been set. Thus the decision maker is induced to make statements about the relative importance of items without knowing what, in fact, is being compared (see section 6.2.10).
- (d) Introducing new options can change the relative ranking of some of the original options. This ‘rank reversal’ phenomenon, first reported by Belton and Gear,⁹¹ is alarming and arises from a failure consistently to relate scales of (performance) measurement to their associated weights.
- (e) Although it is a matter of debate among decision analysts, there is a strong view that the underlying axioms on which AHP is based are not sufficiently clear as to be empirically testable.

5.3 Alternatives to AHP

A number of attempts have been made to develop MCA procedures that retain the strengths of AHP while avoiding some of the objections. The focus of these efforts has largely been on finding different ways of eliciting and then synthesising the pairwise comparisons. It is beyond the scope of the manual to go into great detail about these developments

The best known alternative is *REMBRANDT* (see Lootsma,⁹² and Olson⁹³). *REMBRANDT* uses a direct rating system which is on a logarithmic scale to replace the 1 – 9 scale of AHP and exchanges the eigenvector-based synthesis approach for one which is based on use of the geometric mean to identify estimated weights and scores from pairwise comparison matrices.⁹⁴ A more recent alternative is the *MACBETH* procedure, outlined in section 5.6.

⁸⁹ French, S. (1988) *Decision Theory: an Introduction to the Mathematics of Rationality*, Ellis Horwood, Chichester, pp. 359–361.

⁹⁰ Goodwin, P. and Wright, G. (1998) *Decision Analysis for Management Judgement*, second edition, John Wiley, Chichester.

⁹¹ Belton, V. and Gear, T. (1983) ‘On a short-coming in Saaty’s method of analytic hierarchies’, *Omega*, 11, pp.228–30.

⁹² Lootsma, F.A. (1992) *The REMBRANDT System for Multi-criteria Decision Analysis via Pairwise Comparisons or Direct Rating*, Report 92-05, Faculty of Technical Mathematics and Informatics, Delft University of Technology, The Netherlands.

⁹³ Olson, D. (1996) *Decision Aids for Selection Problems*, Springer Verlag, New York.

⁹⁴ For succinct critiques see French, pp. 359–361, and Goodwin and Wright, pp. 394–397.

Appendix 6 Outranking methods

6.1 Introduction

The bulk of the MCA procedures discussed in this manual are based on work by US and UK researchers. There are however other perspectives on MCA, of which Outranking Methods are arguably the most significant. Outranking as a basis for MCA originated in France in the work of Bernard Roy and colleagues in the mid-1960s and has continued to be applied and extended since that time. Belgian and Dutch researchers have been among the most active in this respect. For a summary of the European School of MCA thinking, see Roy and Vanderpooten.⁹⁵

The main reason for limiting the discussion of outranking procedures in this manual is our view that, for non-specialist users, the MCDA model is more transparent, closer in spirit to established procedures in UK government such as CBA and CEA and provides a clearer and more readily justifiable audit trail. Nonetheless, Outranking Methods also have their strengths. Arguably, they encourage more interaction between the decision maker and the model in seeking out good options. They also recognise directly what is sometimes the political reality, that options recording a poor performance on even one dimension are quite likely to prove unacceptable, because of intensity of lobbying on the part of those bearing the brunt of the poor performance for the criterion concerned.

6.2 Basic procedures

The principal outranking methods assume data availability broadly similar to that required for the MCDA model. That is, they require options to be specified, their performance to be assessed on a series of criteria and for weights to be assessed⁹⁶ that express the relative importance of the criteria.

Outranking is a concept originally due to Roy⁹⁷ and may be defined as follows.

Option *A* outranks Option *B* if, given what is understood of the decision maker's preferences, the quality of the evaluation of the options and the context of the problem, there are enough arguments to decide that *A* is at

⁹⁵ Roy, B. and Vanderpooten, D. (1996) "The European School of MCDA: emergence, basic features and current works", *Journal of Multi-Criteria Decision Analysis*, 5, pp.22–37.

⁹⁶ Strictly speaking, the mathematical definition of weights for at least some outranking methods is different from that used in MCDA. However, it is not necessary to explore these differences in an introduction to outranking. See also Appendix 7 on Outranking Methods with Qualitative Data.

⁹⁷ Roy (1974) "Critères multiples et modélisation des préférences: l'apport des relations de surclassement", *Revue d'Economie Politique*, 1.

least as good as B , while there is no overwhelming reason to refute that statement.

Thus outranking is defined fundamentally at the level of pairwise comparison between every pair of options being considered.

Based on this rather general idea, a series of procedures have been developed to operationalise outranking as a way of supporting multi-criteria decision making. Typically, they involve two phases. First, a precise way of determining whether one option outranks another must be specified. Secondly, it is necessary to determine how all the pairwise outranking assessments can be combined to suggest an overall preference ranking among the options.

Vincke⁹⁸ Chapter 5 provides a clear introduction to the best known outranking methods, of which there are several. Here, just one will be described, the *ELECTRE I* method, due to Roy.⁹⁹ *ELECTRE* derives from *Elimination et Choix Traduisant la Realite*.

6.3 ELECTRE I

ELECTRE I is essentially concerned with identifying dominance¹⁰⁰ relations, in a sense similar, but by no means identical, to dominance as discussed in section 5.5.2.1. It seeks to locate a subset of options, E , such that any option not in E is outranked by at least one member of E . The aim is to make E as small as possible and for it to act as a shortlist, within which a good compromise option should be found.

6.3.1 Phase 1: Defining concordance and discordance

The starting point is to define what are termed the concordance and discordance indices. Using the same notation as in the main body of the manual, the concordance index, $c(i,j)$, can be calculated for every ordered pair of options (i,j) simply as the sum of all the weights for those criteria where option i scores at least as highly as option j .

The discordance index, $d(i,j)$, is a little more complex. If option i performs better than option j on all criteria, the discordance index is zero. If not, then for each criterion where j outperforms i , the ratio is calculated between the difference in performance level between j and i and the maximum observed difference in score on the criterion concerned between any pair of options in the set being considered. This ratio (which must lie between zero and one) is the discordance index.

⁹⁸ Vincke, P. (1992) *Multi-criteria Decision-Aid*, John Wiley, Chichester.

⁹⁹ Roy (1968) 'Classement et choix en presence de points de vue multiples (la methode Electre)', *Revue Francaise de d'Informatique et de Recherche Operationnelle*, 8, pp.57–75.

¹⁰⁰ Throughout appendices 6 and 7, 'dominance' is used to refer to the idea of one option outranking another in the sense defined in section 6.3.2 of this appendix. This is not the same interpretation that dominance has elsewhere in this document, but is common usage in published applications of outranking procedures.

Defined in this way, the discordance index is only of real value in later stages of the analysis if criteria are of roughly equal importance. However, it is possible to refine the discordance definition to avoid this difficulty, albeit at the cost of inducing some element of subjective judgement. It is the discordance index that captures the notion of an option's unacceptability if it records an outstandingly poor performance, even on just one dimension.

6.3.2 Phase 2: Combining concordance and discordance

To bring the two sets of $n(n - 1)$ indices together for all n options being considered, the next phase is to define a (relatively large) concordance threshold, c^* , and a (relatively low) discordance threshold, d^* .

An option then outranks another option overall if its concordance index lies above the chosen threshold value and its discordance index lies below the threshold value.

The set of all options that outrank at least one other option and are themselves not outranked contains the promising options for this problem. If the set is too small, perhaps even an empty set, it can be expanded by appropriate changes to the concordance and/or discordance thresholds. Similarly, if the set is too big, it can be made smaller.

6.4 Assessment

The thinking behind outranking methods is quite different from that underlying MCDA. Although they aim to achieve broadly the same outcome, the two are not easy to compare in a way that does justice to both approaches. In essence, the MCDA approach makes relatively strong assumptions about the underlying circumstances of the problem and delivers in a quite formulaic way a ranking of options, which may then need to be considered carefully to avoid overlooking important factors not captured in the model. MCDA has a relatively strong axiomatic basis.

Outranking methods, on the other hand, seek to make fewer assumptions about the nature of the underlying process that produces preferences. It leaves more of the process of finalising choice to the decision-maker through fine-tuning in terms of items like the concordance and discordance thresholds. It recognises the fact that options which record very poor relative performances on particular dimensions may be hard to implement in practice. It is a more interactive process between decision maker and model.

The view taken in this manual is that the MCDA process is better suited to the broad range of public decisions. However, outranking methods, which have been developed some way beyond the basic *ELECTRE I* model described above, certainly have their proponents and allow perspectives on choice to be developed that MCDA models do not.

6.5 Numerical example

It is assumed that there are ten options (A, B, \dots, J) to be compared, using six criteria. The criteria are assessed on numerical scales, and are such that high values are deemed preferable to low ones. Details are contained in Table A6.1. An assessment of the weights given to each of the six criteria is also required as input data. This is taken as:

$$\underline{w} = [0.25 \ 0.10 \ 0.15 \ 0.25 \ 0.05 \ 0.20]$$

Table A6.1 Option impact matrix for concordance analysis

	A	B	C	D	E	F	G	H	I	J
1	6	2	16	10	11	5	16	17	10	5
2	300	450	350	500	380	250	390	400	410	250
3	27	21	27	20	23	31	24	22	16	18
4	18	19	12	12	20	10	18	26	23	21
5	570	400	420	450	400	430	510	380	410	400
6	12	23	18	20	16	18	21	23	20	22

The first steps of a concordance analysis are the identification of concordance and discordance indices for all pairs of options (i, j), which are then brought together in concordance and discordance matrices. As an illustration, consider the pair of options (C, D). The criteria which form a basis for the calculation of concordance and discordance indices are respectively:

- Concordance Index: {1 ; 3}
- Discordance Index: {2 ; 4 ; 5 ; 6}

The concordance index, $c(C, D)$, is $0.25 + 0.15 = 0.40$, the sum of the weights associated with the members of the concordance set. The discordance index, $d(C, D)$, is the maximum element of the set:

$$\left\{ \frac{150}{250}; 0; \frac{30}{190}; \frac{2}{11} \right\}$$

which are the ratios of the amount of discordance between C and D for each criterion indicating discordance, to the maximum discordance between any pair of options for the corresponding criterion. Hence $d(C, D) = 0.60$. All other elements of the concordance and discordance matrices (shown in Tables A6.2 and A6.3) are computed similarly.

Table A6.2 The concordance matrix										
	A	B	C	D	E	F	G	H	I	J
A	–	0.45	0.30	0.45	0.20	0.65	0.20	0.20	0.20	0.55
B	0.55	–	0.55	0.60	0.30	0.55	0.55	0.15	0.45	0.45
C	0.55	0.45	–	0.40	0.65	0.60	0.15	0.20	0.45	0.55

w

	A	B	C	D	E	F	G	H	I	J
D	0.55	0.40	0.35	–	0.35	0.85	0.10	0.15	0.30	0.55
E	0.80	0.65	0.35	0.65	–	0.60	0.25	0.20	0.40	0.50
F	0.35	0.45	0.02	0.15	0.40	–	0.15	0.20	0.20	0.20
G	0.55	0.45	0.60	0.90	0.75	0.85	–	0.20	0.65	0.55
H	0.80	0.65	0.80	0.85	0.80	0.80	0.80	–	0.85	0.95
I	0.80	0.55	0.55	0.25	0.60	0.80	0.35	0.15	–	0.65
J	0.45	0.50	0.45	0.45	0.45	0.45	0.45	0.05	0.35	–

Table A6.3 The discordance matrix										
	A	B	C	D	E	F	G	H	I	J
A	–	1.00	0.67	0.80	0.36	0.55	0.82	1.00	0.73	0.91
B	0.89	–	0.93	0.53	0.60	0.67	0.93	1.00	0.53	0.20
C	0.79	0.45	–	0.60	0.50	0.27	0.47	0.88	0.69	0.56
D	0.63	0.44	0.47	–	0.50	0.73	0.40	0.88	0.69	0.56
E	0.89	0.64	0.33	0.48	–	0.53	0.58	0.64	0.36	0.55
F	0.74	0.80	0.73	1.00	0.63	–	0.95	1.00	0.81	0.69
G	0.32	0.24	0.20	0.44	0.13	0.47	–	0.50	0.31	0.19
H	1.00	0.20	0.33	0.40	0.10	0.60	0.68	–	0.16	0.11
I	0.84	0.33	0.73	0.76	0.47	1.00	0.53	0.47	–	0.18
J	0.89	0.80	0.73	1.00	0.52	0.87	0.73	0.80	0.64	–

Having prepared the concordance and discordance matrices, it is now necessary to start to identify patterns of dominance among the options, using concordance and discordance thresholds. Sometimes, the mean values of the elements in the respective matrices are used as initial thresholds. In the current example this would imply values $c^* = 0.48$ and $d^* = 0.61$. By definition, dominance of option i over option j occurs at given threshold levels if both $c(i,j) \geq c^*$ and $d(i,j) < d^*$ simultaneously. Here, this yields the following initial dominance pattern:

- A dominates F
- B dominates D
- C dominates E, F, J

- D dominates J
- E dominates D, F, J
- G dominates A, C, D, E, F, I, J
- H dominates B, C, D, E, F, I, J
- I dominates B, E, J

Because of the way dominance is defined, dominance relationships are not transitive. A dominant option is, by definition, one which (a) dominates at least one other option; (b) is not itself dominated. Here, only *G* and *H* are dominant.

In order to try to determine a smaller set of dominant options (perhaps containing just one option), the process of dominance identification can be repeated using more demanding concordance and discordance thresholds. In this example, c^* has been increased in 10 per cent steps, and d^* similarly decreased. This pattern of change is essentially arbitrary, however, and, in theory, different dominance patterns will be found if different adjustment schemes are followed. In practice, the dominance pattern does not appear to be particularly sensitive to choice of adjustment schemes, within a reasonable range.

The changing pattern of dominance as c^* and d^* alter is shown in Table A6.4.

Table A6.4 Changing patterns of dominance as c^* and d^* change				
	c^*	d^*	Dominator	Dominated
1	0.53	0.54	B	D
			C	E, F
			E	D, F
			G	A, C, D, E, F, I, J
			H	B, C, D, E, F, I, J
			I	B, E, J
2	0.58	0.49	C	F
			E	D
			G	C, D, E, F, I
			H	B, C, D, E, F, I, J
			I	E, J
3	0.64	0.44	G	D, E, I
			H	B, C, D, E, F, I, J
			I	J
4	0.70	0.40	G	E
			H	C, D, E, F, I, J

Thus, in this example, given the chosen scheme of adjustment for c^* and d^* , the set of dominating options cannot be reduced below two, *G* and *H*.

If a single “optimum” selection is required, recourse must be had to other techniques.

If it is desired to obtain a ranking of all the options in the pool initially considered, two approaches are possible. One is simply to remove the option ultimately selected as ‘optimal’ and to repeat the analysis on a reduced pool of nine options. The ‘optimum’ in the second application of the concordance analysis becomes the second ranked scheme. The process can be repeated to yield a complete ranking of schemes, although the ranking of the lowest ranked options may be rather arbitrary. Alternatively, a fair general impression of the relative merits of the options can be obtained merely from the output of the analysis just undertaken. A rough ordering of options might be:

H G I C E B A D F J

Appendix 7 Outranking methods with qualitative data

7.1 Introduction

To be able to undertake informative MCA studies without requiring precise numerical data inputs is an attractive possibility. Some criteria do not naturally lend themselves to numerical measurement. Sometimes time or other resource constraints restrict the accuracy with which options' performances can be measured.

In this manual, the recommended approach in such circumstances is either to undertake a limited assessment, based on the procedures set out in Chapter 4, or to use expert judgement to estimate subjectively performances on the relevant 0 – 100 scales for the criteria concerned, supplementing the overall evaluation with sensitivity testing.

An alternative approach, however, is to apply one of the MCA procedures specially designed to require no more than qualitative data inputs. Within this range of MCA methods, only one will be outlined here in any detail.

7.2 Qualitative outranking methods

Nijkamp and Van Delft¹⁰¹ and Voogd¹⁰² both suggest procedures for a qualitative outranking analysis. Here, the version due to Nijkamp and Van Delft is explained. It is based on the assumption that performance on each criterion is categorised into one of four categories (●●●●, ●●●, ●● and ●) in descending order of quality. Similarly, criteria importance assessments (weights) are qualitative and restricted to three categories (◆◆◆, ◆◆, ◆) in decreasing order of importance.

First, pairwise comparisons are made for each criterion between all pairs from among the n options being considered, in a similar way to that set out in Appendix 6. Depending on the difference in assessed performance, each comparison might be:

at one extreme a major positive difference (●●●● against ●) coded as +++;
at the other extreme, a major negative difference (● against ●●●●) coded as – – –; or any of the intermediate assessments ++, +, 0. – or – –.

For any one criterion, all the pairwise comparisons may be summarised by a 'skew-symmetric' matrix, with zeros down the leading diagonal.

¹⁰¹ Nijkamp, P. and Van Delft, A. (1977) *Multi-criteria Analysis and Regional Decision Making*, Martinus Nijhoff, Leiden.

¹⁰² Voogd, H. (1983) *Multi-criteria Evaluation for Urban and Regional Planning*, Pion, London.

Departing now from the procedure followed for *ELECTRE I*, three sets of concordance indices are calculated, one for each of the three criterion importance categories (◆◆◆, ◆◆, ◆). These indices, $c^3(i,j)$, $c^2(i,j)$, and $c^1(i,j)$, represent the frequency with which option i is better than option j , for criteria with high (◆◆◆), medium (◆◆) and low (◆) importance. These outputs may be summarised in three $(n \times n)$ concordance matrices, C^3 , C^2 and C^1 .

Separately for each of the concordance matrices into which these three sets of calculations have been set, it is now possible to compute three net total dominance indices. This is achieved by computing for each matrix the difference between the extent to which option i dominates all other options and the extent to which other options dominate option i (effectively the sum of row i minus the sum of column i). These are denoted c^3 , c^2 and c^1 .

An (unweighted) discordance matrix is now calculated for each pair of plans in a similar way. The discordance indices, $d^3(i,j)$, $d^2(i,j)$, and $d^1(i,j)$, are calculated as the frequency with which the outcomes of option i are much worse (— — —), worse (— —) and slightly worse (—) than option j . This information feeds directly into three discordance matrices from which in turn three net discordance dominance indices may be computed, d^3 , d^2 and d^1 in an analogous fashion to the concordance dominance indices previously described.

Final selection is not based on any fully defined procedure, but revolves around an inspection of the net concordance and discordance indices at each of the three levels of criterion importance, seeking an option that exhibits high concordance and low discordance, especially with respect to the more important weight groups.

7.3 Other qualitative methods

A range of other qualitative methods are described in Nijkamp and Van Delft and in Voogd. Additionally, the more recent Nijkamp et al.¹⁰³ provides other examples and a series of case studies. Within the range of qualitative methods, it is interesting that some begin to merge in terms of their general style and data requirements with the ranking-based approximations to the linear additive model described in Appendix 4.

For example, the Qualiflex method¹⁰⁴ is essentially a permutation analysis, aiming at deriving a rank order of options that is as far as possible consistent with ordinal information contained in the performance matrix and the weight vector. The Regime Analysis method¹⁰⁵ can be interpreted as an ordinal generalisation of pairwise comparison methods such as concordance analysis. Both these methods are set out and illustrated in Nijkamp et al.¹⁰⁶

¹⁰³ Nijkamp, P., Rietveld, T., and Voogd, H. (1990) *Multi-criteria Evaluation in Physical Planning* North Holland, Amsterdam.

¹⁰⁴ Paelinck, J.H. P. (1976) 'Qualitative multiple criteria analysis, environmental protection and multiregional development, *Papers of the Regional Science Association*', 36, pp.59–74; Paelinck, J.H. P. (1977) 'Qualitative multiple criteria analysis: an application to airport location', *Environment and Planning*, 9, pp.883–95; and Van der Linden, J and Stijnen, H. (1995) *Qualiflex version 2.3*, Kluwer, 1995.

¹⁰⁵ Hinlopen, E., Nijkamp, P. and Rietveld, P. (1983) The regime method: a new multi-criteria technique, in P. Hansen (ed.) *Essays and Surveys on Multiple Criteria Decision Making*, Springer Verlag, Berlin.

¹⁰⁶ Nijkamp, P., Rietveld, T., and Voogd, H. (1990) *Multi-criteria Evaluation in Physical Planning* North Holland, Amsterdam.

Appendix 8 Fuzzy MCA

Fuzzy MCA methods are at the moment largely confined to the academic literature or to experimental applications, although ideas about MCA based on fuzzy sets have been discussed for more than twenty years. Fuzzy sets, conceptualised by Zadeh in the 1960s, are broadly equivalent to the sets found in conventional mathematics and probability theory with one important exception. The exception is that, instead of membership of a set being crisp (that is, an element is either definitely a member of a given set or it is not), set membership is graduated, or fuzzy or imprecise.

Set membership is defined by a membership function, $\mu(x)$, taking values between zero and one. Thus a particular issue might be regarded as a member of the set of major social concerns with a membership value of 0.8. A membership function value of 0 conveys definitely not a member of the set, while $\mu = 1$ conveys definitely a member of the set. $\mu = 0.8$ suggests quite a strong degree of belief that the problem is a major one, but not complete certainty.

Proponents of fuzzy MCA would argue that one of the strengths of the fuzzy approach is that it recognises the reality that many of the concepts involved in decision making are far from clear or precise to those involved. Fuzzy sets provide an explicit way of representing that vagueness in the decision maker's mind in an explicit way. Developing this line of argument has led to many suggestions for fuzzy extensions to conventional MCA methods, such as fuzzy outranking methods and fuzzy utility theory.¹⁰⁷ Nonetheless, it remains that, overall in the MCA community, enthusiasm for fuzzy MCA remains muted. Reasons for this include:

- a lack of convincing arguments that the imprecision captured through fuzzy sets and the mathematical operations that can be carried out on them actually match the real fuzziness of perceptions that humans typically exhibit in relation to the components of decision problems;¹⁰⁸
- doubts as to whether **prescriptively** trying to model imprecision, which is in some sense a **descriptive** reflection of the failings of unaided human decision processing, is the right way to provide support to deliver better decisions;
- failure to establish ways of calibrating membership functions and manipulating fuzzy values that have a transparent rationale from the point of view of non-specialists.

In combination, issues such as these continue to throw substantial doubt on the practical value of fuzzy MCA as a practical tool for supporting the main body of public decisions.

¹⁰⁷ Examples can be found, e.g., in Chen, S.J. and Hwang, C.L. (1992) Fuzzy Multiple Attribute Decision Making: Methods and Applications, Springer Verlag, Berlin.

¹⁰⁸ French, S. (1988) Decision Theory: an Introduction to the Mathematics of Rationality, Ellis Horwood, Chichester.

Appendix 9 Using MCA to explore stakeholder viewpoints

9.1 Introduction

The three case studies examined elsewhere in this volume all have as a main goal helping the responsible decision making group to identify one or more of the options being considered as 'the best'. Although it has been emphasised that MCA supports decision makers and certainly does not 'make the decision' in any mechanistic way, the underlying goal is to provide prescriptive guidance.

For some public policy questions, however, diversity of views and other political, social and scientific factors surrounding an issue may make it unrealistic, or at least premature, to expect MCA to serve this function. Nonetheless, the underlying MCA approach may still be used to generate a better understanding of the reasons for divergences in view and to provide a 'map' of the question under debate. It may even suggest ways forward that may be mutually acceptable to stakeholder groups whose fundamental viewpoints are opposed to one another.

The work by Stirling and Mayer, which utilises a number of features of MCA, illustrates some of these points. It is concerned with genetically modified (GM) crops, specifically the use of GM herbicide tolerance and other strategies for the cultivation of oilseed rape. The aim of the study, however, is not to identify a single strategy option. Rather, it uses the MCA structure as a tool for mapping the debate relating to the risks potentially surrounding introduction of such crops. Unlike some other risk assessment procedures, the process used doesn't focus directly on risks, but seeks to throw light on them indirectly by examining the perceived relative performance of the options and the underlying causes of the differences in view.

This work provides an interesting (and contrasting) potential application of the methods described elsewhere in the manual. It represents quite a different mode of application of MCDA which is being used as an exploratory tool, not as any kind of decision support. It has potential to throw light on some questions (for example exploiting the quantitative basis of the approach in order to facilitate the identification of robust alternatives that aren't optimal for any group of stakeholders but are broadly acceptable to most) where cognitive mapping and the like might not be so effective. Rather unusually for an MCA, the study devotes considerable attention to qualitative discussion of the 'framing assumptions' adopted in criteria definition and scoring.

While the simple linear additive model depicted elsewhere in this manual is in general very robust, using it in circumstances where there may be unwillingness to trade-off between attributes or where there might be strong non-linearities in preferences (including thresholds) will stretch the robustness properties considerably.

9.2 Methodology

The study uses what it terms Multi-Criteria Mapping (MCM) to make a comparative appraisal of GM herbicide tolerant oilseed rape with other technical and policy options. A panel of twelve people, all knowledgeable about the issue in question but representing a wide range of perspectives, was asked to undertake a simplified MCDA application to assess the relative performance of six basic options determined by the research team (organic, integrated pest management, conventional and GM strategies under different regulatory regimes), plus up to six others that they themselves were free to introduce.

The analysis essentially worked through the eight steps of the MCDA process set out in Figure 6.1. A key difference, however, relative to standard decision analysis, is that each individual was guided through his or her own *individual* analysis in a separate 2–3 hour session. Thus, from the outset, the aim was not to achieve a consensus view within the group of twelve on the relative attractiveness of the options, but to expose the variety of views and to try to understand better where the differences were most marked and why.

Step 1: Establish the decision context

The decision context was established through the prior contact to arrange the MCDA interview, pre-circulated written material, and then by direct discussion between the researchers and the person concerned. Participants were chosen based on their knowledge of the subject area and in order to reflect a wide range of perspectives. Clearly it was necessary to explain carefully the purpose of the exercise, the nature and role of the MCDA application, and to cover confidentiality and other matters highlighted by the sensitivity of this particular debate.

Step 2: Identify the options to be appraised

These consisted of the six basic options plus any others identified by the interviewee. Nine of the twelve participants added a total of seventeen options.

Step 3: Identify objectives and criteria

Participants were asked to identify up to twelve criteria. Perhaps because each MCDA application was a separate, individual one, little emphasis seems to have been placed on the explicit identification of objectives to which the criteria related. A basic check for preferential independence was undertaken.

Some grouping of criteria was done for later analysis, but not directly to support the MCDA modelling process itself.

Step 4: Scoring

Subjects had some difficulty with this part of the analysis, possibly because of the scientific complexity and high levels of uncertainty surrounding some of the criteria and the scope for context-dependent variability. As a means of exploring uncertainty and variability, participants were asked to state both a 'high' and a 'low' estimate for each criterion score.

Step 5: Weighting

Swing weighting was not employed in this particular study. Rather, participants generally adopted a more intuitive, ad hoc approach. However, they were made aware of the interdependence between weights and scores, so that weights were, in principle, related to the degree of difference between the best and worst performance levels considered for each criterion, as is required for the MCDA model.

Step 6: Combine the weights and scores

This was done using the conventional linear additive model described in Appendix 4.

Step 7: Examine the results

See below

Step 8: Sensitivity analysis

See below.

9.3 Findings

Because the aim of the MCM was not primarily the conventional identification of good options, the analysis and interpretation of results departed from what would normally have been done in steps 7 and 8 of a standard MCDA. The principal analyses that were undertaken for this particular application concerned: the types of criteria nominated by the participants; various sensitivity tests on option rankings; and a consideration of the desirability of developing diverse portfolios of options, rather than the 'single best solution' which is implicit in the basic implementation of MCDA.

Regarding the first of these, it was found that most participants identified criteria that could be categorised under one of five main headings: environment, agriculture, health, economic and social. Selections tended to reflect participants' own professional interests. However, many of these criteria lay outside the scope of official risk assessments.

Building in part on this analysis, a second aspect of the study was a form of sensitivity testing, to explore what were the key factors determining participants' ranking of alternatives. The study explored whether ranking changed significantly depending upon whether optimistic or pessimistic values of the scores were used and in the light of substantial changes in the order of magnitude of individuals' weight assessments. The eliciting of 'optimistic' and 'pessimistic' scores, and the documentation of associated assumptions and conditions, played a crucial role beyond the quantitative assessment of performance. It gave a handle on the key variables in different perspectives on performance itself. Interestingly, the general finding was that there was relatively little change in ranking from either of these sensitivity checks.

However, what was very influential in deciding what options ranked highly was the choice and definition of criteria themselves. In other words, how participants 'framed' the problem, the overall view of the world they brought to the analysis as reflected in the mix of criteria they felt to be important, was critical. This result is quite consistent with the view of many experienced decision analysts, who would argue that time spent determining the criteria in any MCA is the most important time of all, and generally much more so than excessive fine-tuning of the numerical detail of the models themselves. The authors also discuss the implications for the practice of aggregating the criteria of different participants into an overarching value tree.

Thus one of the principal results (that different frames on decision making as reflected in attribute choice are more likely to differentiate alternatives than different weights and/or scores in a shared model framework) is reasonably well established already in the decision research literature. In this respect, it might be asked what precisely further applications of this approach to other topics might achieve. Perhaps the main benefit would be simply a demonstration effect to those responsible for decision making in the area, that stakeholders' framing of problems differs significantly and that this in turn directly underlies the differences that are observed in preferences for alternatives.

In terms of the conduct of risk analyses and, indeed any form of multi-criteria assessment, this finding supports strongly the view that it is important to avoid premature focusing of the analysis around a single (often technical) perspective on option choice. Conventional risk assessments, because of the limited range of criteria they examine, may well be seen as flawed and unacceptable by some stakeholders. Further, the type of approach developed in MCM can also often aid the identification of robust alternatives, options that may not be the very best in any one interest group's eyes, but which perform tolerably well across all the main perspectives that may be brought to bear on a problem. For example, here, an organic option was a robust performer, whereas the status quo option of continuing intensive agriculture (without GM) often ranked poorly.

Finally, the MCM analysis also provided a basis (described in detail in the report) to explore the attraction of selecting not a single option (often the

unnecessarily restrictive way in which many policy choices are presented) but a diverse portfolio embracing some mixture of the options under consideration.

9.4 Assessment

It is important to recognise that the way in which the MCDA pattern of thinking is used in MCM is different in important ways from the prescriptive mode of application set out elsewhere in this manual. It does not aim to identify promising options judged against criteria derived from an agreed set of objectives that reflect the values of the decision making group, or of society as a whole. Rather, its aim is to map the diversity of perspectives that may be taken on an issue, to highlight the key features underlying the differences and to provide a framework for debate. Both in terms of its identification of criteria and of the way in which the individual steps of analysis are undertaken, it does not match what would be done in a conventional MCDA application. That is not its intention.

Provided that a clear distinction is maintained such that the outcomes of an MCM are not interpreted as providing the basis for direct decision support, MCM does represent a useful extension of the basic MCDA pattern of thinking that allows it to throw light on a new set of issues. There are many other mapping procedures that allow problems to be described in formal terms and then debated (e.g., Eden and Ackermann, 1998) or decision making procedures to be analysed (Hodgkinson et al., 1999). MCM, however, has the attraction that it uses a process that will now be broadly familiar to readers of this manual and, because of its structured and quantitative form, allows aspects of problems to be explored that other mapping procedures do not.

9.5 References

Eden, C. and Ackermann, F. (1998) *Making Strategy: the Journey of Strategic Management*, Sage Publishers.

Hodgkinson, G.P., Bown, N.J., Maule, A.J., Glaister, K.W. and Pearman, A.D (1999) Breaking the frame: an analysis of strategic cognition and decision making under uncertainty, *Strategic Management Journal*, vol. 20, pp.977–85.

Stirling, A., and Mayer, S., (1999) *Rethinking Risk: A pilot multi-criteria mapping of a genetically modified crop in agricultural systems in the UK*, (SPRU Report No 21), Science Policy Research Unit.

Appendix 10 Probabilities to preferences

How can probabilities be used in additive multi-criteria models? One approach is to include a ‘confidence’ criterion, assess probabilities and convert them to preference scores (or ‘expected values’)¹⁰⁹ using a ‘proper scoring rule’. Rules that relate scores to probabilities are considered ‘proper’ only if they encourage an assessor to report probabilities that are accurate representations of the assessor’s state of uncertainty.

Probability of success	Score if success	Score if failure
0	$-\infty$	0
0.01s	–150	–2
0.10s	–90	–7
0.20s	–64	–13
0.30s	–48	–20
0.40s	–37	–28
0.50s	–28	–37
0.60s	–20	–48
0.70s	–13	–64
0.80s	–7	–90
0.90s	–2	–150
1.00	0	$-\infty$

The box shows one such proper scoring rule; it is a (roughly) logarithmic rule (i.e., the score/value for probability p is proportional to $\log_{10}p$). For example, if you judge there is an 85 per cent chance of rain tomorrow and therefore a 15 per cent chance of none, then if it does rain, your score is –7, but if it doesn’t rain your score will be –90. The objective, over many probability assessments, is to achieve a score as close to zero as possible. One way to do this is to be very knowledgeable. If you judge the probability to be high and this turns out to be correct, then you are given a low penalty score. But whatever your state of uncertainty, you can also minimise the penalty score by expressing your uncertainty accurately when reporting it as a probability.

On that last point, suppose you decide to cheat a little and say you are 95 per cent sure, hoping for a score of only –2. Of course if it doesn’t rain, then you will get a score of –150. Since your true belief is 85 per cent, you must think there is an 85 per cent chance of earning that score of –2, and a 15 per cent chance of receiving –150. That gives an expected score of:

$$0.85 \times (-2) + 0.15 \times (-150) = -24.2.$$

¹⁰⁹ A modern discussion of ‘The utility of a probability distribution’ can be found in Bernardo, J.M. and Smith, A.F.M., (1994) *Bayesian Theory*, John Wiley & Sons, Chichester, section 2.7.2, pp 69–75.

But the expected score would be closer to zero if you were to report accurately:

$$0.85 \times (-7) + 0.15 \times (-90) = -19.45$$

The expected score associated with an inaccurate report of your uncertainty is always worse than the expected score for your real belief. Surprisingly, this statement is true whatever your real belief. So in the long run the score will be minimised by reporting accurately.

Close examination of the scoring system will show that while the score might be minimised by giving a high probability, you could receive a large negative score if the event does not occur. Being absolutely sure brings a penalty of minus infinity if you are wrong (as someone once said, When betting on a sure thing, always save the bus fare home). So you will have to balance your confidence that the event will occur against your uncertainty that it won't. The scoring system helps to show what that balancing act means. It also provides an audit trail enabling different assessors to be compared after many assessments have been made. The person with the lowest total score is both knowledgeable about the events in question, and is good at reporting their uncertainty accurately.

Because this scoring system is logarithmic, it ensures that the score for a compound event can be derived by summing the scores associated with the component events. For example, if probabilities of 0.6 and 0.5 are assigned to two events, giving scores of -20 and -28 respectively, the probability of both events is $0.6 \times 0.5 = 0.3$, which is associated with a score of -48 , the sum of the two component scores. Other combinations don't work out quite so neatly as this one, but they are as close as can be expected with a system that works with categories of probabilities.

Much more could be said about proper scoring systems for there is a considerable theoretical and empirical literature devoted to the topic, mostly aimed at finding ways of helping people to learn how to assess unbiased probabilities. Rather than extend this brief introduction, here are some common questions about scoring systems and answers to them.

1 Why not use a linear rule, which is simpler and doesn't provide such a big penalty for low probabilities?

There are three reasons, one theoretical, one empirical and one practical. First, a linear rule, which gives a score for p proportional to $1 - p$, is not a proper scoring rule. To take the previous example, if you really are 85 per cent sure, then under a linear rule you will receive a penalty of only -15 if right and a penalty of -85 if wrong. That gives an expected penalty of

$$0.85 \times (-15) + 0.15 \times (-85) = -25.50.$$

Now consider the expected score if you give a more extreme probability, 1.0 (which gives a linear penalty of 0 if correct, and -100 if wrong):

$$0.85 \times (0) + 0.15 \times (-100) = -15.00$$

Obviously the expected penalty is less for reporting a more extreme probability than you believe. Under a linear scoring system, the expected score is always lowest by reporting 1.0 when you are more than 50 per cent sure and 0 when you are less than 50 per cent sure. The linear score is improper because it does not encourage accurate reporting of uncertainty. That is exactly what has happened in empirical tests of the linear rule; people in the experiments gradually drifted in their probability assessments to 0 and 1.0. The practical reason against the linear rule is that most people do not feel the same increase in preference in moving from a probability of 0.1 to 0.2 as from 0.8 to 0.9. The former is usually seen as a bigger difference in preference, and that larger difference is captured in the logarithmic rule.

2 Why not shorten the scale so the maximum penalty is –100?

This could be done; any linear transformation of a proper scoring rule is itself proper. If all the scores were divided by 1.5, then the biggest would be 100 (apart from the score of infinity). But because the scores are given as whole numbers, this shortening of the scale would have introduced linear portions due to rounding errors. Note that the first difference in scores is 2, the next is 5, then 6, 7, 8 and 9. A shorter scale would make some of those differences equal, introducing a linear portion to the scale which might encourage biased reporting of uncertainty.

3 The large penalties associated with low probabilities seem excessive. Couldn't another proper scoring rule be used which is not so extreme?

Certainly the quadratic rule, in which the score associated with p is a function of $(1 - p)^2$, is less penalising of small probabilities. But for this rule the score for a compound event is not the sum of the scores for the component events. Probabilities must multiply and the corresponding penalty scores must add. For this property to hold, the logarithmic score is needed. With the logarithmic scoring system, one way to lessen the effect of the penalties is to reduce the weight in the MCDA model on that criterion. This simulates less concern for risk.

4 Why use the penalty score associated with the probability of success; why not use the expected penalty score?

Because the expected penalty score for p is the same as for $1 - p$. Take the example again. We saw that the expected penalty score for being 85 per cent sure is –19.45. Here is the expected penalty score for 15 per cent:

$$0.15 \times (-90) + 0.85 \times (-7) = -19.45$$

That is the same as for 85 per cent. So if only the expected penalty were used, there would be no point in ever giving a probability less than 0.50.

Glossary

Analytic Hierarchy Process (AHP). An MCA approach invented by mathematician Thomas Saaty, a professor at the University of Pittsburgh. Inputs to an AHP analysis requires paired comparisons of all options separately for each criterion, although users are often encouraged to express comparisons through a set of verbal statements of relative strength. The comparisons take the form of 'How many times more is option A than option B on this criterion?' In addition, paired comparisons are made of all the options with each other. These comparisons of options with options, and criteria with criteria, are then subjected to matrix mathematics which yields numbers called 'weights' (not the same as weights in MCDA) assigned to both the options and the criteria. These weights purport to show the priorities of the options and the criteria, but their interpretation has been the subject of much debate.

Appraisal. The process of examining options and assessing their relative merits. In this guide, and usually in UK central government, it is used to describe analysis before implementation.

Audit trail. In a non-accounting sense: evidence in the form of references, data or documents that enables an investigator to trace the path of past actions or decisions.

Compensatory MCA techniques. Methods for combining assessments on separate criteria into an overall assessment such that lesser scores for an option on some criteria can be offset by greater scores on other criteria, i.e., trade-offs are modelled. Simple weighted averaging models are compensatory. Lexicographic models are not.

Consequence table. See *performance matrix*.

Contingent valuation. A method used to imply valuations, most notably in the environmental field, by asking individuals about their willingness to pay to reduce adverse consequences, such as increased levels of noise or of air pollution, or their willingness to accept sums of money to put up with such consequences.

Cost-benefit analysis (CBA). A term used to describe analysis which examines options and assesses their relative merits by quantifying in monetary terms as many costs and benefits as possible, including items for which the market does not provide a satisfactory measure of value. The basis of the monetary quantification is usually willingness to accept or pay compensation for gains or losses.

Cost-effectiveness analysis (CEA). A term used to describe analysis which examines options which provide the same, or similar, benefits, and which assesses their relative merits by quantifying and comparing the costs of providing them. These costs may include those for which the market does not supply a satisfactory measure of value.

Criterion. One of a number of measures against which options are assessed and compared in a *multi-criteria analysis* for the degree to which they achieve objectives.

Decision analysis/decision theory. Decision analysis and decision theory refer in this manual to any decision aiding approach that is based on expected utility theory and its later extension to decisions with multiple objectives. The theory assumes only that

the decision maker wishes to be consistent in his or her preferences and decisions (see *expected utility theory*). Decision analysis is the applied technology that was developed from decision theory in the 1960s by Professor Howard Raiffa and his colleagues at Harvard University and by Professor Ronald Howard and his colleagues at Stanford University. The theory was extended in 1976 by Ralph Keeney and Howard Raiffa to include decisions with multiple objectives. This latter approach is commonly referred to as multi-criteria decision analysis (MCDA).

Decision tree. A diagram that shows the outcomes that may occur for a series of interdependent decisions sequenced over time. The actual outcome of each of the individual decisions at each stage is not known with certainty. Appropriate analysis of the tree allows the decision maker to develop, from the outset of the decision process, a contingent decision strategy. This indicates what is the best choice to make at each stage in the decision sequence, contingent upon the pattern of earlier decisions and outcomes.

Discounting. The process of comparing benefits and/or costs which are expected to accrue or be incurred at different points in time, by using a discount rate which reflects the decision maker's relative valuation of benefits at different points in time.

Dominance. The situation where, in a pairwise comparison of two options, the first scores higher than the second on at least one criterion and no lower on any of the others. In this case the first option is said to dominate the second.

Evaluation. The process of examining options and assessing their relative merits. In this guide, and usually in UK central government, it used to describe analysis after implementation. The terms 'policy evaluation' and 'post-project evaluation' are often used to describe evaluation in those two areas. In general usage the word evaluation is often used to describe either before or after analysis. It is also often used to describe the process of deciding where the performance of an option with regard to a particular criterion should be placed on a scale in an MCA analysis.

Expected utility theory. The foundation of decision theory. Starting with several basic assumptions about what is meant by coherent (internally consistent) preferences, the theory shows that two elements are logically implied: probabilities expressing degree of belief, and utilities representing subjective value and risk attitude. The theory also shows how those elements should combine to provide a guide to decision making: weighting the utilities by the probabilities for all anticipated consequences of a course of action, then summing those products to give an expected (weighted average) utility. The course of action with the highest expected utility should be chosen.

Facilitated workshops. A small collection of people who share a goal and perform various tasks, with the help of an impartial individual who facilitates the accomplishment of the group's tasks. One form of facilitated workshop is *decision conferencing*, a two- or three-day event involving a work group of key players who wish to address important issues of concern to their organisation, with the help of an outside facilitator and some computer modelling of participants' judgements about the issues. The computer modelling often takes the form of MCDA.

Fuzzy set theory. This approach, originated by Professor Lotfi Zadeh, models the imprecision and ambiguity which people feel about classifying objects by using

numbers from 0 to 1.0 to reflect the 'degree of membership' in a set. For example, a tall, thin person might be assigned the number 0.8 as a member of the imprecise set of 'tall people', and 0.6 in the set of 'thin people'. It would then be possible to say that the person is taller than he or she is thinner. The set membership functions make it possible to combine numbers across many sets, and so one aspect of fuzzy set theory has been the development of fuzzy MCA methods.

Hedonic price techniques. Techniques to infer valuations by using market prices which reflect a range of different criteria. Hedonic house price indices are used to assess valuations of environmental effects by a statistical analysis of all the different factors influencing property prices, so as to identify the impact of specific environmental effects, such as aircraft noise. Hedonic wage equations are used to assess the impact of risk of loss of life from all the other factors influencing wage levels in different occupations.

Interval scale. A scale whose zero-point and unit of measurement are arbitrary. Fahrenheit and Celsius temperature are examples. Any fixed difference on an interval scale represents the same difference in the quality being measured as any other fixed difference. Thus, the difference in temperature represented by a change from 0°C to 50°C is the same temperature difference as a change from 50°C to 100°C, and that statement would be true even if the measurement were made in Fahrenheit. However, because the zero point is arbitrary, ratios of the numbers themselves do not represent meaningful ratios of the temperatures. Thus, ratio statements about the quality being measured are inadmissible when measurements are made on interval scales, whereas ratios of differences in the numbers do represent ratios of differences in the measured quality. See *ratio scale*.

Lexicographic methods. A general approach to the ordering of preferences in which options are compared on the most important criterion, and the best option is chosen unless other options tie for first place. In that case, evaluations on the second most important criterion are considered to break the tie. If that is not possible then the third most important criterion is consulted, and so on until one option can be chosen. There are several variations on this approach which require more than the minimal information of a strict lexicographic method.

Multi-criteria analysis (MCA). Multi-criteria analysis can be used to describe any structured approach to determine overall preferences among alternative options, where the options accomplish several objectives. It is often used in government, as in this manual, to describe those methods which do not rely predominantly on monetary valuations.

Mutual independence of preferences. The case in which scores assigned to options under one criterion are not affected by the scores assigned under another criterion. This condition is weaker than statistical independence, where if event E is statistically independent from F, then F has to be independent of E. Mutual independence of preferences needs to be checked in both directions. For example, in choosing the best meal option from a menu, the relative preferences for main dishes are usually independent of the diner's preferences for wine, but the relative preferences for wine are often not independent of the preferences for main dishes, so mutual independence fails. Mutual preference independence can hold even when options are

correlated in their measures on real-world criteria provided that the criteria express separate aspects of value.

Non-compensatory MCA techniques. Any method, such as the lexicographic rule, that does not permit trade-offs between criteria, i.e., poor values for an option on one criterion cannot be offset by better values on another criterion.

Normative decision models. These are based on fundamental assumptions (axioms) about rational human behaviour, and use mathematical logic to develop ways to rank options that are demonstrably consistent with the underlying axioms. Thus if the axioms are accepted as true, the model provides a potentially indisputable way to rank options.

Objectives. The purposes which an organisation wishes to achieve in areas of concern. Broad overall objectives, or ultimate objectives, are broken into lower-level or intermediate objectives which are more concrete, and these may be further detailed as sub-objectives, immediate objectives, or criteria which are more operational.

Options. Ways of achieving objectives. Options might be policies, programmes, projects, schemes, systems, or anything else about which a decision is needed.

Performance matrix. A matrix or table setting out the performance of each option according to each of the criteria by which options are to be judged. Sometimes referred to as a *consequence table*.

Preference scale. A scale representing relative strength of preference. Relative scales are defined using the available options as anchors: 100 is usually associated with the most preferred option on a given criterion, and 0 is associated with the least preferred option on that criterion. Fixed scales are defined independently of the available options: 100 may be defined as the 'maximum feasible' and 0 as the 'minimum acceptable'. These are both interval scales, so 0 does not mean no preference or no benefit, any more than 0°Celsius means no temperature. In some cases, ratio scales are used. In this case the zero point is not arbitrary; it represents zero cost or no benefit. Only the unit of measurement is arbitrary, and can be defined by establishing a referent for 100, usually the most preferred option or the maximum feasible.

Prescriptive decision models. Prescriptive decision models are practical tools designed to help decision makers make more rational choices. They recognise the limited effectiveness of unaided, intuitive decision making processes, and seek to develop methods that will take decision makers closer to some rational ideal.

Rank reversal. This arises when adding an option to the list of options under consideration causes certain MCA methods, notably the Analytic Hierarchy Process, to reverse the preference ranking between two other, independent alternatives. This is generally regarded as inconsistent with rational decision making.

Ratio scale. A scale whose unit of measurement is arbitrary, but whose zero point is not—it represents zero amount of the quality being measured. Weight, mass and length are examples. If object A measures 100 cm long and object B measures 50 cm,

then A is twice as long as B, and that statement is true whether the measurement is made in inches, centimetres, hands, or any other standard measuring device. Scales used in MCA often do not exhibit this property; for examples, see also *interval scale*.

Shadow prices. Estimates of the costs of resources which represent their true opportunity costs, in circumstances when observed market prices do not. In perfect markets, shadow prices will simply be equal to market prices, but distortions in the market, such as the presence of monopoly power or of taxes which do not correct externalities, lead to a divergence between market prices and shadow prices.

Stated preference. A method to value benefits or costs for which market prices do not exist, which involves implying underlying valuations from individuals' answers to questions about the choices they would make between different hypothetical alternatives. The term stated preference is often used with regard to choices in the transport sector which imply valuations of different types of travel time.

Swing weighting. The method of eliciting relative weights on different criteria. Swing weighting requires judgements of the swing in preference from 0 to 100 on one preference scale as compared to the 0-to-100 swing on another preference scale. The judgements are made by considering the difference between the 0 and 100 positions, and how much that difference matters. Those two considerations take account of the range of real-world difference in the options on the criteria, and the importance of that difference to achieving the overall objective. Swing weighting results in ratio-scale numbers that reflect the relative importance of the criteria.

Bibliography

General reading

Clemen, R. T. (1996). *Making Hard Decisions; An Introduction to Decision Analysis*. (2nd ed.). Duxbury Press, Belmont, CA.

A comprehensive graduate-level textbook on decision analysis that includes many case studies and examples. Chapters 13 to 16 are devoted to modelling preferences.

Decision Conferencing website www.decision-conferencing.com

Contains information on the process and a collection of mainly UK case studies.

French, S. (1988) *Decision Theory: an Introduction to the Mathematics of Rationality*, Ellis Horwood, Chichester.

A rigorous treatment of the mathematical foundations of decision analysis. Comprehensive and clearly written, but mainly of relevance to people needing an in-depth understanding of the principles underlying MCA.

Edwards, W. and Newman, J.R. (1982) *Multiattribute Evaluation*, Sage Publications, Beverly Hills.

The first text to set out the SMART procedure. Although SMART is now usually implemented in a slightly different way, the insights about the application process as a whole are still very valuable.

French, S. (1989) *Readings in Decision Analysis* Chapman and Hall, London.

Chapter 2 is an excellent exposition of decision analysis. The book reprints several case studies, and Professor French adds many wise comments and observations.

Goodwin, P. and Wright, G. (1998) *Decision Analysis for Management Judgement*, second edition, John Wiley, Chichester.

A clearly written text book introducing a range of topics relevant to managerial decision making. Chapter 2 introduces MCDA for appraisal and evaluation, while Chapter 12 extends MCDA to issues of resource allocation and negotiation.

Hammond, J. S., Keeney, R. L., and Raiffa, H. (1999). *Smart Choices: A Practical Guide to Making Better Decisions*, Harvard University Press, Boston, MA.

The jacket description is accurate: '*Smart Choices* offers the first straightforward, easy-to-follow process designed to improve the way you make business decisions, family decisions, personal decisions-any decision that can help you reach your goals.' Combines a non-mathematical version of MCDA with decision tree modelling of uncertainty. Clever stuff written by leaders in the field.

Decision Analysis and Decision Making www.infoharvest.com/infoharv/hotlinks.htm

While not specifically devoted to MCA, this site contains links to web sites and pages related to Decision Analysis and Decision Making.

Journal of Multi-criteria Decision Analysis (John Wiley & Sons, Ltd).

This is a key source on new developments.

Keeney, R.L. (1992) *Value-Focused Thinking: a Path to Creative Decision Making*, Harvard University Press, Cambridge, Massachusetts.

A generally non-technical treatment of MCA. Strong emphasis on application. Pays particular attention to the role underlying values (criteria) should have in stimulating the creation of options, as well as for guiding choice between them.

Keeney, R.L. and Raiffa, H. (1976) *Decisions with Multiple Objectives: Preferences and Value Trade-offs*, Wiley, New York. Reprinted, Cambridge University Press, 1993.

A classic text. An introduction to the subject, but a relatively rigorous one. Following Raiffa's early work, the first text to set out how the von Neumann/ Morgenstern/ Savage subjective expected utility theory might be applied as prescriptive decision support. Insights about application as well as theory.

Olson, D. (1995) *Decision Aids for Selection Problems*, Springer-Verlag, New York.

A relatively recent survey text. It provides a short description of several of the MCA procedures covered in this manual, along with brief numerical examples.

Public Involvement Programme website www.pip.org.uk

A collaborative project developing new ways of involving the public in decision making, from the UK's Institute for Public Policy Research.

Von Winterfeldt, D. and Edwards, W. (1986) *Decision Analysis and Behavioural Research*, Cambridge University Press, Cambridge.

Devotes more attention to decision making uncertainty and some psychological perspectives on decision making, but nonetheless contains several interesting chapters of relevance to the ideas covered in this manual.

Yoon, K.P. and Hwang, C-L. (1995) *Multiple Attribute Decision Making*, Sage, Beverly Hills.

An elementary introduction to a range of MCA procedures. Some of the methods discussed are of doubtful value in practice.

Watson, S.R. and Buede, D.M. (1987) *Decision Synthesis*, Cambridge University Press, Cambridge.

The Theory section is wide-ranging and addresses a range of decision making questions. The Practice section also covers a wide range of material. Oriented towards

larger and more complex applications, but contains practical insights that are of broad interest.

Papers and books relating to applications

Bana e Costa, C., Nunes da Silva, F., & Vansnick, J.-C. (2000). *Conflict dissolution in the public sector: a case study*. European Journal of Operational Research.

Reports the use of MCDA to develop a compromise solution for a new railway link to the port of Lisbon, Portugal. The solution took into account the different value systems of the three stakeholders in the project, who were unable to agree on previously-proposed solutions.

Beinat, E., & Nijkamp, P. (Eds.). (1998). *Multi-Criteria Evaluation in Land-Use Management*. Dordrecht: Kluwer Academic Publishers.

A collection of edited papers showing how MCDA can be and has been applied to issues of land-use management.

Belton, V. (1993) 'Project planning and prioritisation in the social services – an OR contribution', *Journal of the Operational Research Society*, 44, pp. 115–24.

An MCDA application, but the emphasis in the paper is on the process behind setting up the MCA model and the ways in which the process turned out to be of benefit to the client.

Belton, V., Ackermann, F. and Shepherd, I. (1997) 'Integrated support from problem structuring through to alternative evaluation using COPE and VISA', *Journal of Multi-Criteria Decision Analysis*, 6, pp.115–30.

Application in an NHS hospital trust. Uses a computer-based mapping procedure to establish the initial problem context, followed by the application of MCDA.

Brooks, D.G. and Kirkwood, C.W. (1988) 'Decision analysis to select a microcomputer networking strategy: a procedure and a case study', *Journal of the Operational Research Society*, 39, pp.23–32.

An application of multiattribute utility analysis. However, the discussion of the basic model structuring process followed is for the most part equally relevant to the more straightforward MCDA thinking described in this manual.

Brownlow, S.A. and Watson, S.R. (1987) 'Structuring multi-attribute value hierarchies', *Journal of the Operational Research Society*, 38, pp.309–17.

Good introduction to the process of setting up a value tree for a complex MCA problem. The application context is radioactive waste transport.

Butterworth, N.J. (1989) 'Giving up 'The Smoke': a major institution investigates alternatives to being sited in the city', *Journal of the Operational Research Society*, 40, pp.711–17.

A straightforward and clearly described MCDA application, relating to a decision about where to relocate the Bank of England Registrar's Department.

Corner, J.L. and Kirkwood, C.W. (1991) 'Decision analysis applications on the operations research literature, 1970–1988', *Operations Research*, 39, pp.206–19.

A large survey, structured both in terms of application field and methodological issues addressed. Covers a wider range of decision analysis applications than those directly addressed in this manual, but a good starting point, especially if searching from an applications standpoint.

DETR (1998a) *A New Deal for Trunk Roads in England: Understanding the New Approach to Appraisal*,. Available from Highways Economics and Traffic Appraisal Division, DETR, Great Minster House, Zone 3/08, 76 Marsham Street, London SW1P 4DR.

An introduction to the New Approach and to the construction of Appraisal Summary Tables (ASTs) for the 67 schemes in the Roads Review and Stonehenge. An example of MCA used without full weights and trade-offs.

DETR (1998b) *A New Deal for Trunk Roads in England: Guidance on the New Approach to Appraisal*.

Detailed guidance on application. Available at www.communities.gov.uk/itwp/appraisal/guidance

Edwards, W., and Barron, F. H. (1994) 'SMARTS and SMARTER: Improved simple methods for multiattribute utility measurement', *Organizational Behavior and Human Decision Processes*, 60, 306–325.

A revision of the SMART procedure first described by Edwards and Newman. These new methods are still intended for rapid, simple applications of MCDA.

French, S. (1996) 'Multi-attribute decision support in the event of a nuclear accident', *Journal of Multi-Criteria Decision Analysis*, 5, pp.39–57.

Application of HIVIEW and EQUITY to the design of emergency response procedures following a nuclear accident.

Gregory, R., Using stakeholder values to make smarter environmental decisions, *Environment*, 42(5), June 2000.

The Tillamook Bay National Estuary Project is a model of innovative decision making that allows participants to articulate their values, identify alternatives, and make tradeoffs necessary for broadly acceptable agreements.

Hall, P. (1986) 'Managing change and gaining corporate commitment,' *ICL Technical Journal*, 7, 213–27.

Gives an example of how MCDA was used in a decision conference to resolve disagreement between managers and technical staff in ICL concerning a project that

was over budget and short of funds. The MCDA provided a way of finding a new solution that was not a compromise, yet achieved the objectives of both sides.

Hamalainen, R.P. and Leikola, O. (1996) 'Spontaneous decision conferencing with top-level politicians', *OR Insight*, 9(1), pp.24–28.

General, largely non-technical description of an MCA application run with Finnish members of Parliament.

Hamalainen, R.P. and Karjalainen, R (1992) 'Decision support for risk analysis in energy policy', *European Journal of Operational Research*, 56, pp.172–83.

Applies a simple MCDA structure in studying the societal and environmental effects and risks of different energy production alternatives.

Holbourn, M.(1998) 'Decision conferencing-a tool for budget allocation', *Focus on Police Research and Development*, May 1998 (10), 22–3.

Briefly describes how MCDA was applied force-wide by Bedfordshire Police in 1997 for the 1998 financial year.

Jones, M., Hope, C. and Hughes, R. (1990) 'A multi-attribute value model for the study of UK energy policy', *Journal of the Operational Research Society*, 41, pp.919–29.

Application of an MCDA model to the evaluation of UK energy policy options.

Keeney, R.L. (1980) *Siting Energy Facilities*, Academic Press, New York.

The techniques employed go beyond those covered in this manual, but contain useful information on the process of setting up decision analysis models in this area of application.

Keeney, R.L. (1982) 'Decision analysis: an overview', *Operations Research*, 30, pp.803–37.

A succinct review of multiattribute decision analysis, including models that go beyond the range covered in this manual. Guidance on how to apply decision analysis and references to early applications.

Keeney, R. L. (1999). 'Public values and public policy', in J. Shanteau, B. A. Mellers, & D. A. Schum (Eds.), *Decision Science and Technology: Reflections on the Contributions of Ward Edwards* (pp. 291–312). Boston/Dordrecht/London: Kluwer Academic Publishers.

An excellent discussion of the complex interrelationship of public values and public policy from the perspective of a decision analyst experienced in applying all aspects of decision analysis, including MCDA, to public policy issues. He discusses how public values can be obtained and provides seven brief illustrations in the public sector ranging from air pollution control in Southern California, through integrated resource planning in British Columbia, to energy policy in Germany.

Kirkwood, C.W. (1982) 'A case history of a nuclear power plant site selection', *Journal of the Operational Research Society*, 33, pp.353–63.

Involves both uncertainty and multiple attributes. Good discussion of criterion selection and definition, as well as the MCA process as a whole.

Marples, C. and Robertson, G. (1993) 'Option review with HIVIEW', *OR Insight*, 6(3), pp.13–8.

Applies a simple MCDA model to questions concerning the future development of information systems in an organisation.

Pearman, A.D., Mackie, P.J., May A.D and Simon, D. (1989) 'The use of multi-criteria techniques to rank highway investment proposals', in A.G. Lockett and G. Islei (eds.) *Improving Decision Making in Organisations*, Springer Verlag, Berlin, pp. 157–65.

Outlines how a simple MCDA model can be used to help UK local authorities prioritise small-scale highway improvement proposals in a cost-effective way.

Pearman, A.D., Watson, S.M., Tsamboulas, D. and Beuthe, M. (1997) 'Developing appraisal procedures for trans-European networks', in *Transportation Planning Methods*, 1, (Proceedings of the 25th European Transport Forum Annual Meeting, Seminar E), pp.175–86.

Early output from an EU research project developing a simple MCDA model integrated with cost-benefit analysis for strategic level prioritisation of major transport schemes.

Phillips, L. D. (1989) 'People-centred group decision support'. In G. Doukidis, F. Land, & G. Miller (Eds.), *Knowledge-based Management Support Systems* Ellis Horwood, Chichester.

Describes decision conferencing and reports a real case from the private sector in which two MCDA models were constructed with a group in two days, an appraisal model on the first day and a resource model on the second.

Pratt, R., and Belton, V. (1995). *The DRA Programme of Work to Support the LOCS Pre-Feasibility Study (LOCSPREF) Framework for the Assessment of Solution Options* (Unpublished DRA Report), DERA.

An MOD application of MCDA.

Stirling, A., and Mayer, S., (1999) *Rethinking Risk: A pilot multi-criteria mapping of a genetically modified crop in agricultural systems in the UK*, (SPRU Report No 21), Science Policy Research Unit.

Applies multi-criteria mapping, a method for the appraisal of technological risks which combines the audit trail associated with quantitative risk assessment methods with the open-ended framing and qualitative flexibility of participatory deliberative procedures.

Weaver, M., and Morgan, T. (1993) 'Budget making-is there a better way?' *Public Finance and Accountancy* (11 June 1993).

Explains the MCDA process Dudley Metropolitan District Council used to reduce spending for 1993/4 by £17.5 million. The approach, developed by the Strategic Consultancy Services in ICL's Local Government and Health Division, has now been applied in over 50 Local Authorities in the UK.

Software

Many of the most important benefits from applying the MCA procedures described in this manual do not depend upon sophisticated computer packages. In most applications, by far the most valuable benefit is the identification of a relevant set of criteria against which to judge options. From this follows the creation of an explicit framework to allow all options to be comprehensively evaluated in a consistent way. Typically, this need not demand more than simple pencil and paper.

However, once the analysis moves into the later stages of the MCA process and begins to implement compensatory model structures with explicit weighting and scaling, there are undoubtedly advantages to using some computer support. The main benefits are (a) easy amendment to input data (including sensitivity testing) and (b) attractive and informative presentation of outputs.

In the first instance, a lot of progress can be made simply by undertaking the basic linear additive model calculations in a spreadsheet. The toaster evaluation in Chapter 6 was implemented very quickly using a spreadsheet, for example. This facilitates:

- correcting of errors in initial scoring and weighting
- the calculations
- changes to the underlying model structure through addition or deletion of criteria
- sensitivity testing
- use of multiple weight profiles where, e.g., different interest groups may be exhibiting markedly different views about the relative importance of criteria; and
- graphing of results.

Beyond this level of support, there are a number of specially developed computer packages that support the application of the basic MCDA model. The extra benefits include:

- tailored input screens for information on option performance measures
- alternative ways of inputting weight information
- direct on-screen representation of the value tree hierarchy
- some automation of sensitivity testing
- attractive and informative presentation of output information about options' relative performance; and
- opportunity to see directly the effect on options' relative standing of changing assumptions about input data (real-time sensitivity testing).

Below, a number of the better known packages that can implement the MCDA model are described in outline. A useful annual listing of decision analysis software

(including packages to support some decision types outside the range covered in this manual) has in recent years been published in *OR/MS Today*, the bimonthly journal of the United States Institute of Operations Research and Management Science.¹¹⁰

HIVIEW

This package was originally created over twenty years ago at Decisions and Designs, Inc. With support from ICL, the London School of Economics continued the development in the 1980s, producing a DOS version that could be used easily in a group setting. The current Windows version is marketed by Catalyze Ltd.

HIVIEW is the software used to support the case studies set out in Chapter 7. It has the capacity to solve large and complex MCDA problems. It allows the value tree to be both visually created and edited. A variety of graphical input and output displays is available including visual support for data input (e.g., Figure 7.2), comparisons of options by the importance of their weighted criteria (Figure 7.11) and efficiency frontier presentation of overall costs and benefits (Figure 7.12). It also provides a mechanism for sensitivity analysis to test robustness (Figure 7.20). Input data can be exported to a spreadsheet for further modelling, then imported back into HIVIEW.

MACBETH

MACBETH supports the process of taking possibly incomplete qualitative judgements about the difference in attractiveness of pairs of options and converting them into numerical scores. These scores are entirely consistent with the qualitative pairwise judgements. The MACBETH approach can also be applied to determining criteria weights. MACBETH is particularly useful in public sector applications when a new program, MULTI-MACBETH provides MCDA modelling along with the MACBETH scoring and weighting approach.

The authors are Carlos Bana e Costa, Jean-Marie De Corte and Jean-Claude Vansnick. For further information, contact Carlos Bana e Costa at cbana@alfa.ist.utl.pt

VISA

VISA is another Windows-based implementation of the basic MCDA model. It is marketed by Visual Thinking and has been developed at Strathclyde University.

Its functionality is broadly similar to that of HIVIEW. It, too, permits on-screen creation and editing of the value tree and provides similar input and output display possibilities.

¹¹⁰Buede, D.M. (1998) 'Software survey: decision analysis', *OR/MS Today*, 25(4), pp.56–64.

DESYSION DESKTOP

Desysion Desktop supports application of the MCDA model in a Windows environment.

Developed by the company DecideWise International BV in Amsterdam, it implements MCDA in a way that places special emphasis on guiding decision makers through the whole of the overall process of decision making.

KRYSALIS SOFTWARE SUITE

After converting Hiview and Equity to Windows, Krysalis developed its own set of MCDA tools, **OnBalance**, **HiPriority** and **NoRegrets**. These are the first three in a set of linked packages that share the same modules, file structure and HCI, thus enabling decision makers to build initial models in one package and then view the results from another perspective.

OnBalance is used to evaluate and select alternatives when multiple, conflicting objectives obscure decision making. It allows multiple value trees within the same model, limits to provide reference points for scoring, weighting, analysing, improved sensitivity functions/ displays and multiple mapping types that allow textual and non-linear transformations.

HiPriority is used mainly for resource allocation, project prioritisation and design to best value. It allows three types of interactions between the options/projects, which model dependencies, synergies and alternatives.

NoRegrets allows multiple sets of users to express rich value systems to show different perspectives on a proposed deal. The system will search for equal trades for two or more parties to support the process of moving towards agreement – or having clarity about why a deal is not attractive.

OTHER PACKAGES

Other packages that can provide support to implement the basic MCDA model include the Logical Decisions Package and HIPRE 3+. The latter supports the implementation of a number of different MCA support procedures, including both basic MCDA and AHP. Apart from HIPRE 3+, support for AHP implementation is available also through the Expert Choice package.

Details of the suppliers of these packages are provided below. A number will provide limited trial demonstration versions, many of which can be downloaded from their websites. In addition, HIPRE 3+ is accessible and can be freely used over the worldwide web.

VENDOR DETAILS

Catalyze Ltd.
Yew Tree Cottage
Main Road
Colden Common
Winchester
SO21 1RR
Phone: 01962 715286
Fax: 0870 855 5277
Email: info@catalyze.co.uk
www.catalyze.co.uk

DecideWise International BV
Paasheuvelweg 15
1105 BE Amsterdam
The Netherlands
Phone: +31-20-691-2699
Fax: +31-20-691-1902
e-mail: info@decidewise.com
www.decidewise.com

Expert Choice Inc.
5001 Baum Blvd. Ste. 560
Pittsburgh,
PA 15213
USA
www.expertchoice.com

Krysalis
Phone: +44 (0) 1628 636 861
e-mail: info@krysalis.co.uk
www.krysalis.co.uk

Logical Decisions
1014 Wood Lily Drive
Golden
CO 80401
USA
Phone: +1-303-526-7536
Fax: +1-303-526-7537
e-mail: gary@logicaldecisions.com
www.logicaldecisions.com

Systems Analysis Laboratory
Helsinki University of Technology
Otakaari 1 M
02150 Espoo
Finland
Phone: +358-9-451-3054
Fax: +358-9-451-3096
e-mail: raimo@hut.fi
www.hipre.hut.fi

Visual Thinking International Ltd
141 St James Road
Glasgow
G4 0LT
UK
Phone: 0141 552 6888
Fax: 0141 303 8388
e-mail: info@VisualT.com
www.visualt.com

