EVALUATION AND DECISION MODELS:

a critical perspective

EVALUATION AND DECISION MODELS:

a critical perspective

Denis Bouyssou
ESSEC

■

Thierry Marchant
Ghent University

■

Marc Pirlot
SMRO, Faculté Polytechnique de Mons

■

Patrice Perny
LIP6, Université Paris VI

■

Alexis Tsoukiàs
LAMSADE - CNRS, Université Paris Dauphine

■

Philippe Vincke
SMG - ISRO, Université Libre de Bruxelles

KLUWER ACADEMIC PUBLISHERS
Boston/London/Dordrecht

Contents

1 Introduction **1**
1.1 Motivations 1
1.2 Audience 2
1.3 Structure 3
1.4 Outline 3
1.5 Who are the authors ? 5
1.6 Conventions 5
1.7 Acknowledgements 6

2 Choosing on the basis of several opinions **7**
2.1 Analysis of some voting systems 8
2.1.1 Uninominal election 9
2.1.2 Election by rankings 13
2.1.3 Some theoretical results 16
2.2 Modelling the preferences of a voter 18
2.2.1 Rankings 18
2.2.2 Fuzzy relations 21
2.2.3 Other models 23
2.3 The voting process 23
2.3.1 Definition of the set of candidates 23
2.3.2 Definition of the set of the voters 24
2.3.3 Choice of the aggregation method 24
2.4 Social choice and multiple criteria decision support 25
2.4.1 Analogies 25
2.5 Conclusions 26

3 Building and aggregating evaluations **29**
3.1 Introduction 29
3.1.1 Motivation 29
3.1.2 Evaluating students in Universities 30
3.2 Grading students in a given course 31
3.2.1 What is a grade? 31
3.2.2 The grading process 31
3.2.3 Interpreting grades 37
3.2.4 Why use grades? 40
3.3 Aggregating grades 41
3.3.1 Rules for aggregating grades 41

3.3.2 Aggregating grades using a weighted average 42
3.4 Conclusions 51

4 Constructing measures **53**
4.1 The human development index 54
4.1.1 Scale Normalisation 56
4.1.2 Compensation 57
4.1.3 Dimension independence 58
4.1.4 Scale construction 59
4.1.5 Statistical aspects 59
4.2 Air quality index 61
4.2.1 Monotonicity 61
4.2.2 Non compensation 62
4.2.3 Meaningfulness 62
4.3 The decathlon score 63
4.3.1 Role of the decathlon score 65
4.4 Indicators and multiple criteria decision support 66
4.5 Conclusions 69

5 Assessing competing projects **71**
5.1 Introduction 71
5.2 The principles of CBA 73
5.2.1 Choosing between investment projects in private firms . . . 73
5.2.2 From Corporate Finance to CBA 75
5.2.3 Theoretical foundations 76
5.3 Some examples in transportation studies 79
5.3.1 Prevision of traffic 80
5.3.2 Time gains 80
5.3.3 Security gains 81
5.3.4 Other effects and remarks 82
5.4 Conclusions 83

6 Comparing on several attributes **87**
6.1 Thierry's choice 87
6.1.1 Description of the case 88
6.1.2 Reasoning with preferences 91
6.2 The weighted sum 97
6.2.1 Transforming the evaluations 98
6.2.2 Using the weighted sum on the case 99
6.2.3 Is the resulting ranking reliable? 99
6.2.4 The difficulties of a proper usage of the weighted sum . . . 101
6.2.5 Conclusion 105
6.3 The additive value model 106
6.3.1 Direct methods for determining single-attribute value functions 107

6.3.2 AHP and Saaty's eigenvalue method 111
6.3.3 An indirect method for assessing single-attribute value functions and trade-offs 117
6.3.4 Conclusion 124
6.4 Outranking methods 124
6.4.1 Condorcet-like procedures in decision analysis 124
6.4.2 A simple outranking method 129
6.4.3 Using ELECTRE I on the case 131
6.4.4 Main features and problems of elementary outranking approaches 139
6.4.5 Advanced outranking methods: from thresholding towards valued relations 141
6.5 General conclusion 144

7 Deciding automatically **147**
7.1 Introduction 147
7.2 A System with Explicit Decision Rules 149
7.2.1 Designing a decision system for automatic watering 150
7.2.2 Linking symbolic and numerical representations 150
7.2.3 Interpreting input labels as scalars 153
7.2.4 Interpreting input labels as intervals 156
7.2.5 Interpreting input labels as fuzzy intervals 161
7.2.6 Interpreting output labels as (fuzzy) intervals 164
7.3 A System with Implicit Decision Rules 170
7.3.1 Controlling the quality of biscuits during baking 170
7.3.2 Automatising human decisions by learning from examples . 171
7.4 An hybrid approach for automatic decision-making 174
7.5 Conclusion 176

8 Dealing with uncertainty **179**
8.1 Introduction 179
8.2 The context 179
8.3 The model 180
8.3.1 The set of actions 180
8.3.2 The set of criteria 181
8.3.3 Uncertainties and scenarios 182
8.3.4 The temporal dimension 184
8.3.5 Summary of the model 186
8.4 A didactic example 186
8.4.1 The expected value approach 187
8.4.2 Some comments on the previous approach 187
8.4.3 The expected utility approach 189
8.4.4 Some comments on the expected utility approach 191
8.4.5 The approach applied in this case: first step 193
8.4.6 Comment on the first step 196
8.4.7 The approach applied in this case: second step 198

8.5 Conclusions . . . 200

9 Supporting decisions **205**

9.1 Preliminaries . . . 206
9.2 The Decision Process . . . 207
9.3 Decision Support . . . 210
9.3.1 Problem Formulation . . . 211
9.3.2 The Evaluation Model . . . 213
9.3.3 The final recommendation . . . 219
9.4 Conclusions . . . 226
Appendix A . . . 228
Appendix B . . . 231

10 Conclusion **237**

10.1 Formal methods are all around us . . . 237
10.2 What have we learned? . . . 239
10.3 What can be expected? . . . 243

Bibliography **247**

Index **262**

6

COMPARING ON THE BASIS OF SEVERAL ATTRIBUTES: THE EXAMPLE OF MULTIPLE CRITERIA DECISION ANALYSIS

6.1 Thierry's choice

How to choose a car is probably the multiple criteria problem example that has been most frequently used to illustrate the virtues and possible pitfalls of multiple criteria decision aiding methods. The main advantage of this example is that the problem is familiar to most of us (except for one of the authors of this book who is definitely opposed to owning a car) and it is especially appealing for male decision-makers and analysts for some psychological reason. However, one can object that in many illustrations, the problem is too roughly stated to be meaningful; the motivations, needs, desires and/or phantasms of the potential buyer of a new or second-hand car can be so diversified that it will be very difficult to establish a list of relevant points of view and build criteria on which everybody would agree; the price for instance is a very delicate criterion since the amount of money the buyer is ready to spend clearly depends on his social condition. The relative importance of the criteria also very much depends on the personal characteristics of the buyer: there are various ideal types of car buyers, for instance people who like sportive car driving, or large comfortable cars or reliable cars or cars that are cheap to run. One point should be made very clear: it is unlikely that a car could be universally recognised as the best, even if one restricts oneself to a segment of the market; this is a consequence of the existence of decision-makers with many different "value systems".

Despite these facts, we have chosen to use the "Choosing a car" example, in a properly defined context, for illustrating the hypotheses underlying various elementary methods for modelling and aggregating evaluations in a decision aiding process. The case is simple enough to allow for a short but complete description; it also offers sufficient potential for reasoning on quite general problems raised by the treatment of multi-dimensional data in view of decision and evaluation. We describe the context of the case below and will invoke it throughout this chapter for illustrating a sample of decision aiding methods.

	Trademark and type
1	Fiat Tipo 20 ie 16V
2	Alfa 33 17 16V
3	Nissan Sunny 20 GTI 16
4	Mazda 323 GRSI
5	Mitsubishi Colt GTI
6	Toyota Corolla GTI 16
7	Honda Civic VTI 16
8	Opel Astra GSI 16
9	Ford Escort RS 2000
10	Renault 19 16S
11	Peugeot 309 GTI 16V
12	Peugeot 309 GTI
13	Mitsubishi Galant GTI 16
14	Renault 21 20 turbo

Table 6.1: List of the cars selected as alternatives

6.1.1 Description of the case

Our example is adapted from an unpublished report by a Belgian engineering student who describes how he decided which car he would buy. The story dates back to 1993; our student—call him Thierry—aged 21, is passionate about sportive cars and driving (he has taken lessons in sports car driving and participates in car races). Being a student, he cannot afford to buy either a new car nor a luxury second hand sports car; so he decides to explore the middle range segment, 4 year old cars with powerful engines. Thierry intends to use the car in everyday life and occasionally in competitions. His strategy is first to select the make and type of the car on the basis of its characteristics, estimated costs and performances, then to look for such a car in second hand car sale advertisements. This is what he actually did, finding "the rare pearl" about twelve months after he made up his mind as to which car he wanted.

Selecting the alternatives

The initial list of alternatives was selected taking an additional feature into account. Thierry lives in town and does not have a garage to park the car in at night. So he does not want a car that would be too attractive to thieves. This explains why he discards cars like VW Golf GTI or Honda CRX. He thus limits his selection of alternatives to the 14 cars listed in Table 6.1.

Selecting the relevant points of view and looking for or constructing indices that reflect the performances of the alternatives for each of the viewpoints often constitutes a long and delicate task; it is moreover a crucial one since the quality of the modelling will determine the relevance of the model as a decision aiding tool. Many authors have advocated a hierarchical approach to criteria building, each viewpoint being decomposed into sub-points that can be further decomposed

(Keeney and Raiffa (1976), Saaty (1980)). A thorough analysis of the properties required of the family of criteria selected in any particular context (*consistent* family, i.e. *exhaustive*, *non-redundant* and *monotonic*) can be found in Roy and Bouyssou (1993) (see also Bouyssou (1990), for a survey).

We shall not emphasise the process of selecting viewpoints in this chapter, although it is a matter of importance. It is sufficient to say that Thierry's concerns are very particular and that he accordingly selected five viewpoints related to cost (criterion 1), performance of the engine (criteria 2 and 3) and safety (criteria 4 and 5).

Evaluations of the cars on these viewpoints have been obtained from monthly journals specialised in the benchmarking of cars. The official quotation of second hand vehicles of various ages is also published in such journals.

Evaluating the alternatives

Evaluating the expenses incurred by buying and using a specific car is not as straightforward as it may seem. Large variations from the estimation may occur due to several uncertainty and risk factors such as actual life-length of the car, actual selling price (in contrast to the official quotation), actual mileage per year, etc. Thierry evaluates the expenses as the sum of an initial fixed cost and expenses resulting from using the car. The fixed costs are the amount paid for buying the car, estimated by the official quotation of the 4-year old vehicle, plus various taxes. The yearly costs involve another tax, insurance and petrol consumption. Maintenance costs are considered roughly independent of the car and hence neglected. Petrol consumption is estimated on the basis of three figures that are highly conventional: the number of litres of petrol burned in 100 km is taken from the magazine benchmarks; Thierry somehow estimates his mileage at 12 000 km per year and the price of the petrol to .9 € per litre (1 €, the European currency unit, is approximately equivalent to 1 USD). Finally he expects (hopes) to use the car for 4 years. On the basis of these hypotheses he gets the estimations of his expenses for using the car during 4 years that are reported in Table 1 (Criterion 1 = Cost). The resale value of the car after 8 years is not taken into account due to the high risk of accidents resulting from Thierry's offensive driving style. Note that the petrol consumption cost which is estimated with a rather high degree of imprecision counts for about one third of the total cost. The purchase cost is also highly uncertain.

For building the other criteria Thierry has a large number of performance indices whose value is to be found in the magazine benchmarks at his disposal. Thierry's particular interest in sporty cars is reflected in his definition of the other criteria. Car performances are evaluated by their acceleration; criterion 2 ("Accel" in Table 6.2) encodes the time (in seconds) needed to cover a distance of one kilometre starting from rest. One could alternatively have taken other indicators such as power of the engine, time needed to reach a speed of 100 km/h or to cover 400 meters that are also widely available. Some of these values may be imprecisely determined: they may be biased when provided by the car manufacturer (the procedures for evaluating petrol consumption are standardised but usually under-

	Name of cars	Crit1 Cost	Crit2 Accel	Crit3 Pick up	Crit4 Brakes	Crit5 Road-h
1	Fiat Tipo	18 342	30.7	37.2	2.33	3
2	Alfa 33	15 335	30.2	41.6	2	2.5
3	Nissan Sunny	16 973	29	34.9	2.66	2.5
4	Mazda 323	15 460	30.4	35.8	1.66	1.5
5	Mitsubishi Colt	15 131	29.7	35.6	1.66	1.75
6	Toyota Corolla	13 841	30.8	36.5	1.33	2
7	Honda Civic	18 971	28	35.6	2.33	2
8	Opel Astra	18 319	28.9	35.3	1.66	2
9	Ford Escort	19 800	29.4	34.7	2	1.75
10	Renault 19	16 966	30	37.7	2.33	3.25
11	Peugeot 309 16V	17 537	28.3	34.8	2.33	2.75
12	Peugeot 309	15 980	29.6	35.3	2.33	2.75
13	Mitsubishi Galant	17 219	30.2	36.9	1.66	1.25
14	Renault 21	21 334	28.9	36.7	2	2.25

Table 6.2: Data of the "choosing a car" problem

estimate the actual consumption for everyday use); when provided by specialised journalists in magazines, the procedures for measuring are generally unspecified and might vary since the cars are not all evaluated by the same person.

The third criterion that Thierry took into consideration is linked with the pick up or suppleness of the engine in urban traffic; this dimension is considered important since Thierry also intends to use his car in normal traffic. The indicator selected to measure this dimension ("Pick up" in Table 6.2) is the time (in seconds) needed for covering one kilometre when starting in fifth gear at 40 km/h. Again other indicators could have been chosen (e.g. the torque). This dimension is not independent of the second criterion, since they are generally positively correlated (powerful engines generally lead to quick response times on both criteria); cars that are specially prepared for competition may however lack suppleness in low operation conditions which is quite unpleasant in urban traffic. So, from the point of view of the user, i.e. in terms of preferences, criteria 2 and 3 reflect different requirements and are thus both necessary. For a short discussion about the notions of independence and interaction, the reader is referred to Section 6.2.4.

In the magazine's evaluation report, several other dimensions are investigated such as comfort, brakes, road-holding behaviour, equipment, body, boot, finish, maintenance, etc. For each of these, a number of aspects are considered: 10 for comfort, 3 for brakes, 4 for road-holding, ... In view of Thierry's particular motivations, only the qualities of braking and of road-holding are of concern to him and lead to the building of criteria 4 and 5 (resp. "Brakes" and "Road-h" in Table 6.2). The 3 or 4 partial aspects of each viewpoint are evaluated on an ordinal scale the levels of which are labelled "serious deficiency", "below average", "average", "above average", "exceptional". To get an overall indicator of braking quality (and also for road-holding), Thierry re-codes the ordinal levels with integers from 0 to 4 and takes the arithmetic mean of the 3 or 4 numbers; this results in the figures with 2 decimals provided in the last two columns of Table 1. Obviously these numbers are also imprecise, not necessarily because of imprecision in the evaluations but because of the arbitrary character of the cardinal re-coding of the ordinal information and its aggregation via an arithmetic mean (postulating implicitly that, in some sense, the 3 components of each viewpoint are equally important and the levels of each of the three scales are equally spaced). We shall however consider that these figures reflect, in some way, the behaviour of each car from the corresponding viewpoint; it is clear however that not too much confidence should be awarded to the precision of these "evaluations".

Note that the first 3 criteria have to be minimised while the last 2 must be maximised.

This completes the description of the "data" which, obviously, are not given but selected and elaborated on the basis of the available information. Being intrinsically part of this data is an appreciation (more or less explicit) of their degree of precision and their reliability.

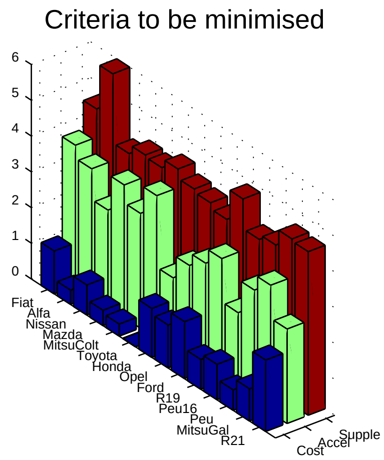
6.1.2 Reasoning with preferences

In the second part of the presentation of this case, Thierry will provide information about his preferences. In fact, in the relatively simple decision situation he was facing ("no wife, no boss", Thierry decides for himself and the consequences of his decision should not affect him crucially), he was able to make up his mind without using any formal aggregation method. Let us follow his reasoning.

First of all he built a graphic representation of the data. Many types of representations can be thought of; popular spreadsheet software offer a large number of graphical options for representing multi-dimensional data. Figure 6.1 shows such a representation. Note that the evaluations for the various criteria have been re-scaled in view of a better readability of the figure. The values for all criteria have been mapped (linearly) onto intervals of length 2, the first criterion being represented in the $[0, 2]$ interval, the second criterion, in the $[2, 4]$ interval and so on. For each criterion, the lowest evaluation observed for the sample of cars is mapped on the lower bound of the interval while the highest value is represented on the upper bound of the interval. Such a transformation of the data is not always innocent; we briefly discuss this point below.

In view of reaching a decision, Thierry first discards the cars whose braking efficiency and road-holding behaviour is definitely unsatisfactory, i.e. car numbers 4, 5, 6, 8, 9, 13. The reason for such an elimination is that a powerful engine is needless in competition if the chassis is not good enough and does not guarantee good road-holding; efficient brakes are also needed to keep the risk inherent to competition at a reasonable level. The rules for discarding the above mentioned cars have not been made explicit by Thierry in terms of unattained levels on the corresponding scales. Rules that would restate the set of remaining cars are for instance:

$$\text{criterion } 4 \geq 2$$



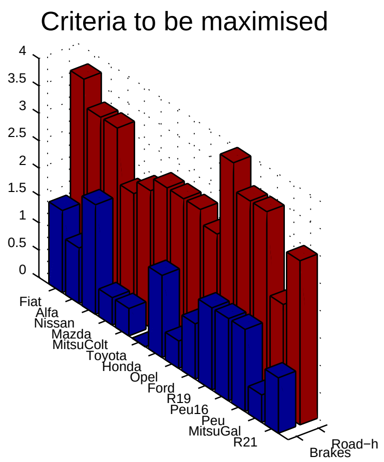


Figure 6.1: Performance diagram of all cars along the first three criteria (above; to be minimised) and the last two (below; to be maximised)

and

$$\text{criterion } 5 \geq 2$$

with at least one strict inequality.

Looking at the performances of the remaining cars, those labelled 1, 2, 10 are further discarded. The set of remaining cars is restated for instance by the rule:

$$\text{criterion } 2 < 30$$

Finally, the car labelled 14 is eliminated since it is *dominated* by car number 11. "Dominated by car 11" means that car 11 is at least as good on all criteria and better on at least one criterion (here all of them!). Notice that car number 14 would not have been dominated if other criteria had been taken into consideration such as comfort or size: this car is indeed bigger and more classy than the other cars in the sample.

The cars left after the above elimination process are those labelled 3,7,11,12; their performances are shown on Figure 6.2. In these star-diagrams each car is represented by a pentagon; their values on each criterion have all been linearly re-scaled, being mapped on the $[1, 3]$ interval. The choice of interval $[1, 3]$ instead of interval $[0, 2]$ is dictated by the mode of representation: the value "0" plays a special role since it is common to all axes; if an alternative was to receive a 0 value on several criteria, those evaluations would all be represented by the origin, which makes the graph less readable. On each axis, the value 1 corresponds to the lowest value for one of the cars in the initial set of 14 alternatives on each criterion; the value 3 corresponds to the highest value for one of the 14 cars. In interpreting the diagrams, remember that criteria 1, 2 and 3 are to be minimised while the others have to be maximised.

Thierry did not use the latter diagram (Figure 6.2); he drew the same diagram as in Figure 6.1 instead after reordering the cars; the 4 candidate cars were all put on the right of the diagram as shown in Figure 6.3; in this way Thierry was still able to compare the difference in the performances of two candidate cars for a criterion to typical differences for that criterion in the initial sample. This suggests that the evaluations of the selected cars should not be transformed independently of the values of the cars in the initial set; these still constitute reference points in relation to which the selected cars are evaluated. On Figure 6.4, for the reader's convenience, we show a close-up of Figure 6.3 that is focused on the 4 selected cars only.

Thierry first eliminates car number 12 on the basis of its relative weakness on the second criterion (acceleration). Among the 3 remaining cars the one he chooses is number 11. Here are the reasons for this decision.

1. Comparing cars 3 and 11, Thierry considers that the price difference (about 500 €) is worth the gain (.7 second) on the acceleration criterion.

2. Comparing cars 7 and 11, he considers that the cost difference (car 7 about 1 500 € more expensive) is not balanced by the small advantage on acceleration (.3 second) coupled with a definite disadvantage (.8 second) on suppleness.

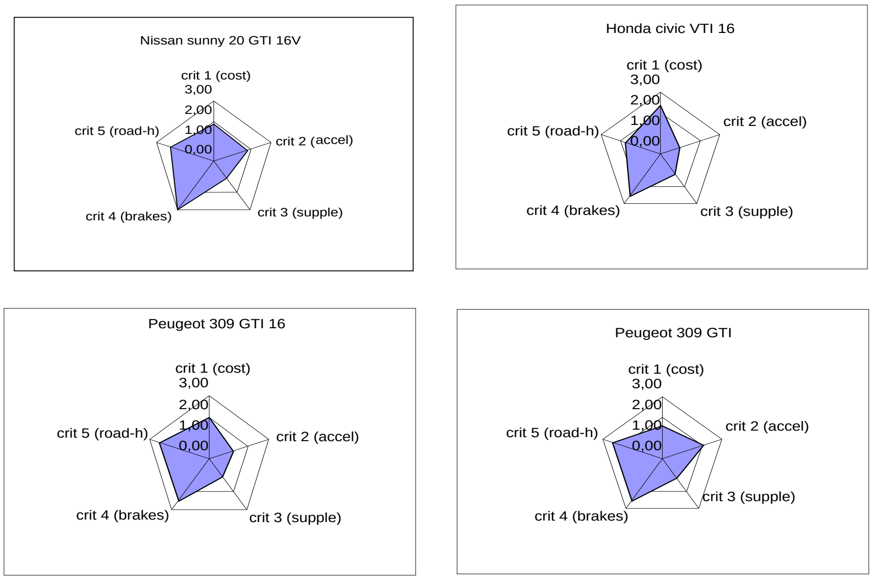


Figure 6.2: Star graph of the performances of the 4 cars left after the elimination process

	Name of car	Crit1 Cost	Crit2 Acc	Crit3 Pick	Crit4 Brakes	Crit5 Road
3	Nissan Sunny	16 973	29	34.9	2.66	2.5
7	Honda Civic	18 971	28	35.6	2.33	2
11	Peugeot 16V	17 537	28.3	34.8	2.33	2.75
12	Peugeot	15 980	29.6	35.3	2.33	2.75

Table 6.3: Performances of the 4 candidate cars

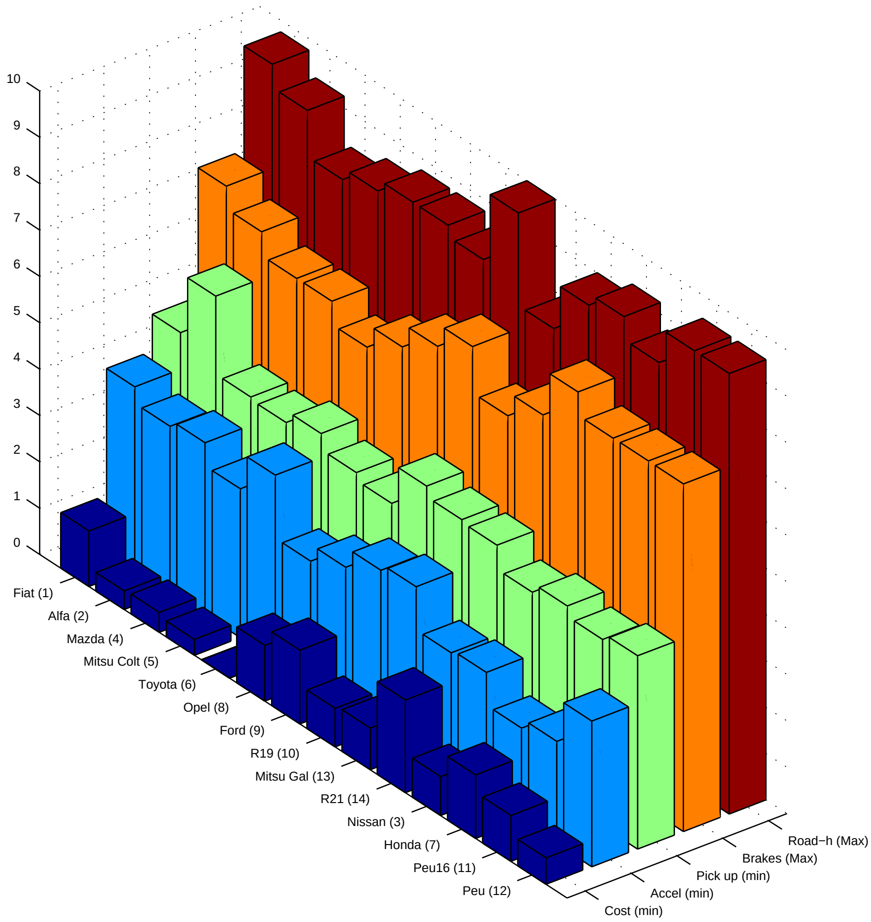


Figure 6.3: Performance diagram of all cars; the 4 candidate cars stand on the right

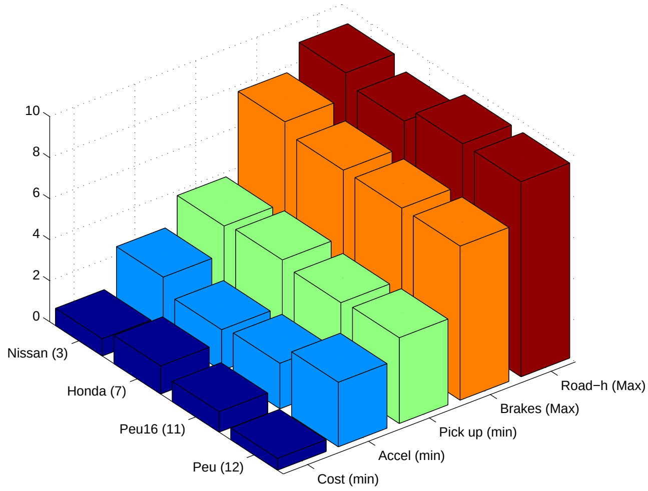


Figure 6.4: Detail of Figure 6.3: the 4 cars remaining after initial screening

Comments

Thierry's reasoning process can be analysed as being composed of two steps. The first one is a *screening* process in which a number of alternatives are discarded on the basis of the fact that they do not reach *aspiration levels* on some criteria.

Notice that these levels have not been set *a priori* as minimal levels of satisfaction; they have been set after having examined the whole set of alternatives, to a value that could be described as both desirable and accessible. The rules that have been used for eliminating certain alternatives have exclusively been combined in *conjunctive* mode since an alternative is discarded as soon as it does not fulfil any of the rules.

More sophisticated modes of combinations may be envisaged, for instance mixing up conjunctive and disjunctive modes with aspiration levels defined for subsets of criteria (see Fishburn (1978) and Roy and Bouyssou (1993), pp. 264-266). Another elementary method that has been used is the elimination of *dominated* alternatives (car 11 dominates car 14).

In the second step of Thierry's reasoning,

1. Criteria 4 and 5 were not invoked; there are several possible reasons for this: criteria 4 and 5 might be of minor importance or considered satisfactory once a certain level is reached; they could be insufficiently discriminating for the considered subset of cars (this is certainly the case for criterion 4): the values of the differences for the set of candidate cars could be such that they are not large enough to balance the differences on other criteria.

2. Subtle considerations on whether the balance of differences in performance between pairs of cars on 2 or 3 criteria results in an advantage to one of the cars in the pair.

3. The reasoning is not made on the basis of re-coded values like those used in the graphics; more intuition is needed, which is better supported by the original scales. Since criteria 4 and 5 are aggregates and, thus, are not expressed in directly interpretable units, this might also have been a reason for not exploiting them in the final selection.

This kind of reasoning that involves comparisons of differences in evaluations is at the heart of the activity of modelling preferences and aggregating them in order to have an informed decision process. In the simple case we are dealing with here, the small number of alternatives and criteria has allowed Thierry to make up his mind without having to build a formal model of his preferences. We have seen, however, that after the first step consisting in the elimination of unsatisfactory alternatives, the analysis of the remaining four cars has been much more delicate.

Note also that if Thierry's goal had been to rank order the cars in order of decreasing preference, it is not sure that the kind of reasoning he used for just choosing the best alternative for him would have fit the bill. In more complex situations (when more alternatives remain after an initial elimination or more criteria have to be considered or if a ranking of the alternatives is wanted), it may appear necessary to use tools for modelling preferences.

There is another rather frequent circumstance in which more formal methods are mandatory; if the decision-maker is bound to justify his decision to other persons (shareholders, colleagues, ...), the evaluation system should be more systematic, for instance being able to cope with new alternatives that could be suggested by the other people.

In the rest of this chapter, we discuss a few formal methods commonly used for aggregating preferences. We report on how Thierry applied some of them to his case and extrapolate on how he could have used the others. This can be viewed as an *ex post* analysis of the problem, since the decision was actually made well before Thierry became aware of multiple criteria methods. In his *ex post* justification study, Thierry has in addition tried to derive a ranking of the alternatives that would reflect his preferences.

6.2 The weighted sum

When dealing with multi-dimensional evaluations of alternatives, the basic and almost natural (or perhaps, cultural?) attitude consists in trying to build a one-dimensional synthesis, which would reflect the value of the alternatives on a synthetic "super scale of evaluation". This attitude is perhaps inherited from school practice where all other performance evaluations of the pupils have long been (and often still are) summarised in a single figure, a weighted average of their grades in the various subjects. The problems raised by such a practice have been discussed in depth in Chapter 3. We discuss the application of the weighted sum to the car example below, emphasising the very strong hypotheses underlying the use of this type of approach.

Starting from the standard situation of a set of alternatives $a \in A$ evaluated on n points of view by a vector $g(a) = (g_1(a), g_2(a), \ldots, g_n(a))$, we consider the value $f(a)$ obtained by linearly combining the components of g , i.e.

$$f(a) = k_1 g_1(a) + k_2 g_2(a) + \ldots + k_n g_n(a) \tag{6.1}$$

Suppose, without loss of generality, that all criteria are to be maximised, i.e. the larger the value $g_i(a)$, the better the alternative a on criterion i (if, on the contrary, g_i were to be minimised, substitute g_i by $-g_i$ or use a negative weight k_i). Once the weights k_i have been determined, choosing an alternative becomes straightforward: the best alternative is the one associated with the largest values of f. Similarly, a ranking of the alternatives is obtained by ordering them in decreasing order of the value of f.

This simple and most commonly used procedure relies however on very strong hypotheses that can seldom be considered plausibly satisfied. These problems appear very clearly when trying to use the weighted sum approach on the car example.

6.2.1 Transforming the evaluations

A look at the evaluations of the cars (see Table 6.2) prompts a remark that was already made when we considered representing the "data" graphically. The ranges of variation on the scales are very heterogeneous: from 13841 to 21334 on the cost criterion; from 1.33 to 2.66 on criterion 4. Clearly, asking for values of the weights k_i in terms of the relative importance of the criteria without referring to the scales would yield absurd results. The usual way out consists in normalising the values on the scales but there are several manners of doing this. One consists in dividing g_i by the largest value on the i^{th} scale, $g_{i,max}$; alternatively one might subtract the minimal value $g_{i,min}$ and divide by the range $g_{i,max} - g_{i,min}$. These normalisations of the original g_i functions are respectively denoted g'_i and g''_i in the following formulae

$$g'_i(a) = \frac{g_i(a)}{g_{i,max}} \tag{6.2}$$

$$g''_i(a) = \frac{g_i(a) - g_{i,min}}{g_{i,max} - g_{i,min}} \tag{6.3}$$

For simplicity, we suppose here that g_i are positive. In the former case the maximal value of g'_i will be 1 while value 0 is kept fixed which means that the ratio of the evaluations of any pair a, b of alternatives remains unaltered:

$$\frac{g'_i(a)}{g'_i(b)} = \frac{g_i(a)}{g_i(b)} \tag{6.4}$$

This transformation can be advanced when using *ratio scales*, in which the value 0 plays a special role. Statements such as "alternative a is twice as good as b on criterion i" remain valid after transformation.

In the case of g_i, the top evaluation will be mapped onto 1 while the bottom one goes onto 0; ratios are not preserved but ratios in differences of evaluations do: for all alternatives a, b, c, d,

$$\frac{g_i''(a) - g_i''(b)}{g_i''(c) - g_i''(d)} = \frac{g_i(a) - g_i(b)}{g_i(c) - g_i(d)} \tag{6.5}$$

Such a transformation is appropriate for *interval scales*; it does not alter the validity of statements like "the difference between a and b on criterion i is twice the difference between c and d".

Note that the above are not the only possible options for transforming the data; note also that these transformations depend on the set of alternatives: considering the 14 cars of the initial sample or the 4 cars retained after the first elimination would yield substantially different results since the values $g_{i,min}$ and $g_{i,max}$ depend on the set of alternatives.

6.2.2 Using the weighted sum on the case

Suppose we consider that 0 plays a special role in all scales and we choose the first transformation option. The values of the g_i's that are obtained are shown in Table 6.4. A set of weights has been chosen which is, to some extent, arbitrary but seems compatible with what is known about Thierry's preferences and priorities. The first three criteria receive negative weights namely and respectively -1, -2, -1 (since they have to be minimised), while the last two are given the weight .5. The alternatives are listed in Table 6.4 in decreasing order of the values of f. As can be seen in the last column of Table 6.4, this rough assignment of weights yields car number 3 as first choice followed immediately by car number 11 which was actually Thierry's choice. Moreover, the difference in the values of f for those two cars is tiny (less than .01) but we have no idea as to whether such a difference is meaningful; all we can do is being very prudent in using such a ranking since the weights were chosen in a rather arbitrary manner. It is likely that by varying the weights slightly from their present value, one would readily get rank reversals i.e. permutations of alternatives in the order of preference; in other words, the ranking is not very stable. Varying the values that are considered imprecisely determined is what is called sensitivity analysis; it helps to detect what the *stable* conclusions in the output of a model are; this is certainly a crucial activity in a decision aiding process.

6.2.3 Is the resulting ranking reliable?

Weights depend on scaling

To illustrate the lack of stability of the ranking obtained, let us consider Table 6.5 where the set of alternatives is reduced to the 4 cars remaining after the elimination procedure; the re-scaling of the criteria yields values of g_i that are not the same as in Table 6.4 since $g_{i,max}$ depends on the set of alternatives. This perturbation, without any change in the values of the weights, is sufficient to cause a rank reversal between the leading two alternatives. Of course, one could prevent such a drawback, by using a normalising constant that would not depend on the

		Weights k_i					Value
		-1	-2	-1	0.5	0.5	f
Nr	Name of cars	Cost	Accel	Pick	Brak	Road	
3	Nissan Sunny	0.80	0.94	0.84	1.00	0.77	-2.63
11	Peugeot 16V	0.82	0.92	0.84	0.88	0.85	-2.64
12	Peugeot	0.75	0.96	0.85	0.88	0.85	-2.66
10	Renault 19	0.80	0.97	0.91	0.88	1.00	-2.71
7	Honda Civic	0.89	0.91	0.86	0.88	0.62	-2.82
1	Fiat Tipo	0.86	1.00	0.89	0.88	0.92	-2.85
5	Mitsu Colt	0.71	0.96	0.86	0.62	0.54	-2.91
2	Alfa 33	0.72	0.98	1.00	0.75	0.77	-2.92
8	Opel Astra	0.86	0.94	0.85	0.62	0.62	-2.96
6	Toyota	0.65	1.00	0.88	0.50	0.62	-2.97
4	Mazda 323	0.72	0.99	0.86	0.62	0.46	-3.02
9	Ford Escort	0.93	0.95	0.83	0.75	0.54	-3.03
14	Renault 21	1.00	0.94	0.88	0.75	0.69	-3.04
13	Mitsu Galant	0.81	0.98	0.89	0.62	0.38	-3.15

Table 6.4: Normalising then ranking through a weighted sum

		Weights k_i					Value
		-1	-2	-1	0.5	0.5	f
Nr	Name of car	Cost	Accel	Pick	Brak	Road	
11	Peugeot 16V	0.92	0.96	0.98	0.88	1.00	-2.876
3	Nissan Sunny	0.89	0.98	0.98	1.00	0.91	-2.890
12	Peugeot	0.84	1.00	0.99	0.88	1.00	-2.896
7	Honda Civic	1.00	0.95	1.00	0.88	0.73	-3.090

Table 6.5: Normalising then ranking a reduced set of alternatives

set of alternatives, for instance the worst acceptable value (minimal requirement for a performance to be maximised; maximal level of a variable to be minimised, a cost, for instance) on each criterion; with such an option, the source of the lack of stability would be the imprecision in the determination of the worst acceptable value. Notice that the above problem has already been discussed in Chapter 4, Section 4.1.1.

Conventional codings

Another comment concerns the figures used for evaluating the performances of the cars on criteria 4 and 5. Recall that those were obtained by averaging equally spaced numerical codings of an ordinal scale of evaluation. The obtained figures presumably convey a less quantitative and more conventional meaning than for instance acceleration performances measured in seconds in standardisable (if not standardised) trials. These figures however are treated in the weighted sum just like the "more quantitative" ones associated with the first three criteria. In particular, other codings of the ordinal scale might have been envisaged, for instance codings with unequal intervals separating the levels on the ordinal scale. Some of these codings could obviously have changed the ranking.

6.2.4 The difficulties of a proper usage of the weighted sum

The meaning of the weights

What is the exact significance of the weights in the weighted sum model? The weights have a very precise and quantitative meaning; they are *trade-offs*: to compensate for a disadvantage of k_i units for criterion j, you need an advantage of k_j units for criterion i. An important consequence is that the weights depend on the determination of the unit on each scale. In a weighted sum model that would directly use the evaluations of the alternatives given in Table 6.2, it is clear that the weight of criterion 2 (acceleration time) has to be multiplied by 60 if times are expressed in minutes instead of seconds. This was implicitly a reason for normalising the evaluations as was done through formulae 6.2 and 6.3. After transformation, both g'_i and g''_i are independent of the choice of a unit; yet they are not identical and, in a consistent model, their weights should be different. Indeed, we have

(6.6) $$g''_i(a) = \frac{g_{i,max}}{g_{i,max} - g_{i,min}} \times g'_i(a) + \lambda_i = \kappa_i \times g'_i(a) + \lambda_i$$

where λ_i is a constant. Additive constants do not matter since they do not alter the rating. So, unless $g_{i,min} = 0$, g''_i is essentially related to g'_i by a multiplicative factor $\kappa_i \neq 1$; in order to model the same preferences through a weighted sum of the g''_i and a weighted sum of the g'_i , the weight k''_i of g''_i should be obtained by dividing the weight k'_i by κ_i. Obviously, the weights have to be assessed in relation to a particular determination of the evaluations on each scale and eliciting them in practice is a complex task. In any case, they certainly cannot be evaluated in a meaningful manner through naive questions about the relative importance of the criteria; reference to the underlying scale is essential.

Up to this point we have considered the influence on the weights of multiplying the evaluations by a positive constant. Note that translating the origin of a scale has no influence on the ranking of the alternatives provided by the weighted sum since it results in adding a (positive or negative) constant to f, the same for all alternatives. There is still a very important observation that has to be made: all scales used in the model are implicitly considered linear in the sense that equal differences in values on a criterion result in equal differences in the overall evaluation function f and this does not depend on the position of the interval of values corresponding to that difference on the scale. For instance in the car example, car number 12 is finally eliminated because it accelerates too slowly. The difference between car 12 and car 3 with respect to acceleration is 0.6 between 29 seconds and 29.6 seconds. Does Thierry perceive this difference as almost equally important as a difference of 0.7 between cars 11 and 3, the latter difference being positioned between 28.3 seconds and 29 seconds on the acceleration scale? It seems rather clear from Thierry's motivations, that coming close to a performance of 28 seconds is what matters to him while cars above 29 seconds are unworthy. This means that the gain for passing from 29.6 seconds to 29 seconds has definitely less value than a gain of similar amplitude, say from 29 to 28.3 seconds. As will be confirmed in the sequel (see Section 6.3 below), it is very unlikely that Thierry's preferences are correctly modelled by a linear function of the current scales of performance.

Independence or interaction

The next issue is more subtle. Evaluations of the alternatives for the various points of view taken into consideration by the decision-maker often show correlations; this is because the attributes that are used to reflect these viewpoints are often linked by logical or factual interdependencies. For instance, indicators of cost, comfort and equipment, which may be used as attributes for assessing the alternatives for those viewpoints, are likely to be positively correlated. This does not mean that the corresponding points of view are redundant and that one should eliminate some of them. One is perfectly entitled to work with attributes that are (even strongly) correlated. That is the first point.

A second point is about independence. In order to use a weighted sum, the viewpoints should be independent, but not in the statistical sense implying that the evaluations of the alternatives should be uncorrelated! They should be independent *with respect to preferences*. In other words, if two alternatives that share the same profile on a subset of criteria compare in a certain way in terms of overall preferences, their relative position should not be altered when the profile they share on a subset of criteria is substituted by any other common profile. On the contrary, a famous example of dependence in the sense of preferences in a gastronomic context is the following: the preference for white wine or red wine usually depends on whether you are eating fish or meat. There are relatively simple tests for independence in the sense of preferences, which consist in asking the decision-maker about his preferences on pairs of alternatives that share the same profile for a subset of attributes; varying the common profile should not reverse the preferences when the points of view are independent. Independence is a necessary condition for the representation of preferences by a weighted sum; it is not a sufficient one of course.

There is a different concept that has been recently implemented for modelling preferences. It is the concept of *interacting* criteria that was already discussed in example 2 of Chapter 3. Suppose that in the process of modelling the preferences of the decision-maker, he declares that the influence of positively correlated aspects should be dimmed and that conjoint good performances for negatively correlated aspects should be emphasised. In our case for instance, criteria 2 and 3, respectively acceleration and suppleness, may be thought of as being positively correlated. It may then prove impossible to model some preferences by means of a weighted sum of the evaluations such as those in Table 6.2 (and even of transformations thereof such as obtained through formulae like 6.3). This does not mean that no additive model would be suitable and it does not imply that the preferences are not independent (in the above-defined sense). In the next section we shall study an additive model, more general than the weighted average, in which the evaluations g_i may be "re-coded" through using "value functions" u_i. With appropriate choices of u_2 and u_3 it may be possible to take the decision-maker's preferences about positively and negatively correlated aspects into account, provided they satisfy the independence property. If no re-coding is allowed (like in the assessment of students, see Chapter 3) there is a non-additive variant of the weighted average that could help modelling interactions among the criteria; in such a model the weight of a coalition of criteria may be larger or smaller than the sum of the weights of its components (see Grabisch (1996), for more detail on non-additive averages).

Arbitrariness, imprecision and uncertainty

In the above discussion as well as in the presentation of our example we have emphasised the many sources of uncertainty (lack of knowledge) and of imprecision that bear on the figures used as input in the weighted sum. Let us summarise some of them:

1. Uncertainty in the evaluation of the cost: the buying price as well as the life-length of a second hand car are not known. This uncertainty can be considered of stochastic nature; statistical data could help to master—to some extent—such a source of uncertainty; in practice, it will generally be very difficult to get sufficient relevant and reliable statistical information in for this kind of problems.

2. Imprecision in the measurement of some quantities: for instance, how precise is the measurement of the acceleration? Such an imprecision can be reduced by making the conditions of the measurement as standard as possible and can then be estimated on the basis of the precision of the measurement apparatus.

3. Arbitrary coding of non-quantitative data: re-coding of ordinal scales of appreciation of braking and road-holding behaviour. Any re-coding that respects the order of the categories would in principle be acceptable. To master such an imprecision one could try to build quantitative indicators for the criteria or try to get additional information on the comparison between differences of levels on the ordinal scale: for instance, is the difference between "below average" and "average" larger than the difference between "above average" and "exceptional"?

4. Imprecision in the determination of the trade-offs (weights k_i); the ratios of weights k_j/k_i must be elicited as conversion rates: a unit for criterion j is worth k_j/k_i units for criterion i; of course, the scales must first be re-coded in order that one unit difference on a criterion has the same "value" everywhere on the scale (linearisation); these operations are far from obvious and as a consequence, the imprecision of the linearisation process combines with the inaccuracy in the determination of weights.

Making a decision

All these sources of imprecision have an effect on the precision of the determination of the value of f that is almost impossible to quantify; contrary to what can (often) be done in physics, there is generally little information on the size of the imprecisions; quite often, there is not even probabilistic information on the accuracy of the evaluations. As a consequence, the apparently straightforward decision—choosing the alternative with the highest value of f or ranking the alternatives in decreasing order of the values of f—might be unconsidered as illustrated above. The usual way out is extensive sensitivity analysis, which could be described as part of the validation of the model. This part of the job is seldom carried out with the required exhaustivity because it is a delicate task at least in two respects. On the one hand there are many possible strategies for varying the values of the imprecisely determined parameters; usually parameters are varied one at a time which is not sufficient but is possibly tractable; the range in which the parameters must be varied is not even clear as suggested above. On the other hand, once the sensitivity analysis has been performed, one is likely to be faced with several almost equally valuable alternatives; in the car problem for instance, the simple remarks made above strongly suggest that it will be very difficult to discriminate between cars 3 and 11.

In view of the previous discussion, there are two main approaches to solve the difficulties raised by the weighted sum:

1. Either one tries to prepare the inputs of the model (linearised evaluations and trade-offs) as carefully as possible, paying permanent attention to reducing imprecision and finishing with extensive sensitivity analysis;

2. Or one takes imprecision into account from the start, by avoiding to exploit precise values when knowing that they are not reliable but rather working with classes of values and ordered categories. Note that imprecision may well

lie in the link between evaluations and preferences rather than in the evaluations themselves; detailed preferential information, even extracted from perfectly precise evaluations, may prove rather difficult to elicit.

The former option will lead us to the construction of *multi-attribute value* or *utility functions*, while the latter leads to the *outranking approach*. These two approaches will be developed in the sequel. There is however a whole family of methods that we shall not consider here, the so-called *interactive methods* (Steuer (1986), Vincke (1992), Teghem (1996)). These implement various strategies for exploring the efficient boundary, i.e. the set of non-dominated solutions; the exploration jumps from one solution to another; it is guided by the decision-maker who is asked to tell, for instance, which characteristics of the current solution he would like to see improved. Such methods are mainly designed for dealing with infinite and even continuous sets of alternatives; moreover, they do not lead to an explicit model of the decision-maker's preferences. On the contrary, we have settled on problems with a (small) finite number of alternatives and we concentrate on obtaining explicit representations of the decision-maker's preferences.

6.2.5 Conclusion

The weighted sum is useful for obtaining a quick and rough draft of an overall evaluation of the alternatives. One should however keep in mind that there are rather restrictive assumptions underlying a proper use of the weighted sum. As a conclusion to this section we summarise these conditions.

1. **Cardinal character of the evaluations on all scales.** The evaluations of the alternatives for all criteria are numbers and these values are used as such even if they result from the re-coding of ordinal data.

2. **Linearity of each scale.** Equal differences between values on scale i, whatever the location of the corresponding intervals on the scale (at the bottom, in the middle or at the top of the scale), produce the same effect on the overall evaluation f: if alternatives a, b, c, d are such that $g_i(a) - g_i(b) = g_i(c) - g_i(d)$ for all i, then $f(a) - f(b) = f(c) - f(d)$.

3. **The weights are trade-offs.** Weights depend on the scaling of the criteria; transforming the (linearised) scales results in a related transformation of the weights. Weights tell how many units on the scale of criterion i are needed to compensate one unit of criterion j.

4. **Preference independence.** Criteria do not interact. This property, called preference independence, can be formulated as follows. Consider two alternatives that share the same evaluation on at least one criterion, say criterion i. Varying the level of that common value on criterion i does not alter the way the two alternatives compare in the overall ranking.

6.3 The additive multi-attribute value model

Our analysis of the weighted sum brought us very close to the requirements for additive multi-attribute value functions. The most common model in multiple criteria decision analysis is a formalisation of the idea that the decision-maker, when making a decision, behaves as if he was trying to maximise a quantity called *utility* or *value* (the term "utility" tends nowadays to be used preferably in the context of decision under risk, but we shall use it sometimes for "value").
This postulates that all alternatives may be evaluated on a single "super-scale" reflecting the value system of the decision-maker and his preferences. In other words, the alternatives can be "measured", in terms of "worth" on a synthetic dimension of value or utility. Accordingly, if we denote by $\succsim$ the overall preference relation of the decision-maker on the set of alternatives, this relation relates to the values $u(a)$, $u(b)$ of the alternatives in the following way:

$$a \succsim b \text{ iff } u(a) \geq u(b) \tag{6.7}$$

As a consequence, the preference relation $\succsim$ on the set of alternatives is a complete preorder, i.e. a complete ranking possibly with ties. Of course, the value $u(a)$ usually is a function of the evaluations $\{g_i(a), i = 1, \ldots, n\}$. If this function is a linear combination of $g_i(a), i = 1, \ldots, n$, we get back to the weighted sum. A slightly more general case is the following additive model:

$$u(a) = \sum_{i=1}^{n} u_i(g_i(a)) \tag{6.8}$$

where the function u_i (single-attribute value function) is used to re-code the original evaluation g_i in order to linearise it in the sense described in the previous section; the weights k_i are incorporated in the u_i functions. The additive value function model can thus be viewed as a clever version of the weighted sum since it allows us to take some of the objections—mainly the second hypothesis in Section 6.2.5—against a naive use of it into account. Note however that the imprecision issue is not dealt with inside the model (sensitivity analysis has to be performed in the validation phase, but is neither part of the model nor straightforward in practice); the elicitation of the partial value functions u_i may also be a difficult task.

Much effort has been devoted to characterising various systems of conditions under which the preferences of a decision-maker can be described by means of an additive value function model. Depending on the context, some systems of conditions may be interpretable and tested, at least partially, i.e. it may be possible to ask the decision-maker questions that will determine whether an additive value model is compatible with what can be perceived of his system of preferences. If the preferences of the decision-maker are compatible with an additive value model, a method of elicitation of the u_i's may then be used; if not, another model should be looked for: a multiplicative model or, more generally, a non-additive one, a non-independent one, a model that takes imprecision more intrinsically into account, etc. (see Krantz, Luce, Suppes and Tversky (1971), Chapter 7, Luce, Krantz, Suppes and Tversky (1990), Vol. 3, Chapter 19).

6.3.1 Direct methods for determining single-attribute value functions

A large number of methods have been proposed to determine the u_i's in an additive value function model. For an accessible account of such methods, the reader is referred to von Winterfeldt and Edwards (1986), Chapter 8.

There are essentially two families of methods, one based on direct numerical estimations and the other on indifference judgements. We briefly describe the application of a technique of the latter category relying on what is called *dual standard sequences*, (Krantz et al. (1971), von Winterfeldt and Edwards (1986), Wakker (1989)) that builds a series of equally spaced intervals on the scale of values.

An assessment method based on indifference judgments

Suppose we want to assess the u_i's in an additive model for the *Cars* case. It is assumed that the suitability of such a model for representing the decision-maker's preferences has been established. Consider a pair of criteria, say *Cost* and *Acceleration*. We are going to outline a simulated dialog between an analyst and a decision-maker that could yield an assessment of u_1 and u_2, the corresponding single-attribute value functions, for ranges of evaluations corresponding to acceptable cars. Note that we start the construction of the sequence from a "central point" instead of taking a "worst point" (see for instance von Winterfeldt and Edwards (1986), pp. 267 *sq* for an example starting from a worst point)

The range for the cost will be the interval between 21 500 € to 13 500 € and from 28 to 31 seconds for acceleration. First ask the decision-maker to select a "central point" corresponding to medium range evaluations on both criteria. In view of the set of alternatives selected by Thierry, let us start with $(17\,500, 29.5)$ as "average" values for cost and acceleration. Also ask the decision-maker to define a unit step on the cost criterion; this step will consist, say, of passing from a cost of 17 500 € to 16 500 €. Then the standard sequence is constructed by asking which value x_1 for the acceleration would make a car costing 16 500 € and accelerating in 29.5 seconds indifferent to a car costing 17 500 € and accelerating in x_1 seconds. Suppose the answer is 29.2 meaning that from the chosen starting point, a gain of 0.3 second on the acceleration time is worth an increase of 1 000 € in cost. The answer could be explained by the fact that at the starting level of performance for the acceleration criterion, the decision-maker is quite interested by a gain in acceleration time. Relativising the gains as percentages of the half range from the central to the best values on each scale, this means that the decision-maker is ready to lose $\frac{1000}{4000}$=25% of the potential reduction in cost for gaining $\frac{.3}{1.5}$=20% of acceleration time. We will say in the sequel that the parity is equal when the decision-maker agrees to exchange a percentage of the half range on a criterion against an equal percentage on another criterion.

The second step in the construction of the standard sequence is asking the decision-maker which value to assign to x_2 to have $(16\,500, 29.2) \sim (17\,500, x_2)$, where $\sim$ denotes "indifferent to". The answer might be, for instance, 28.9. Continuing along the same line would for instance yield the following sequence of

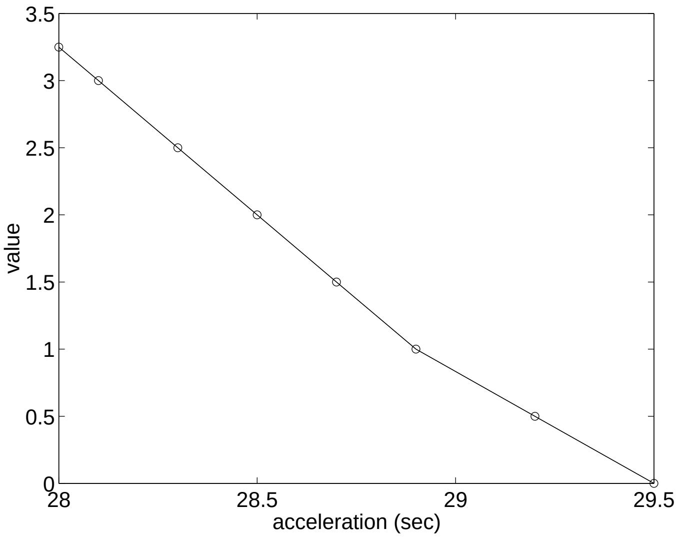


Figure 6.5: Single-attribute value function for acceleration criterion (half range)

indifferences:

$$\begin{array}{ccc}(16\,500, 29.5) & \sim & (17\,500, 29.2)\\ (16\,500, 29.2) & \sim & (17\,500, 28.9)\\ (16\,500, 28.9) & \sim & (17\,500, 28.7)\\ (16\,500, 28.7) & \sim & (17\,500, 28.5)\\ (16\,500, 28.5) & \sim & (17\,500, 28.3)\\ (16\,500, 28.3) & \sim & (17\,500, 28.1)\end{array}$$

Such a sequence gives the analyst an approximation of the single-attribute value function u_2, on the half range from 28 to 29.5 seconds but it is easy to devise a similar procedure for the other half range, from 29.5 to 31. Figure 6.5 shows the re-coding u_2 of the evaluations g_2 on the interval $[28, 29.5]$; there are two linear parts in the graph: one ranging from 28 to 28.9 where the slope is proportional to $\frac{1}{.2}$ and the other valid between 28.9 and 29.5 with a slope proportional to $\frac{1}{.3}$.

From there, using the same idea, one is able to re-code the scale of the cost criterion into the single-attribute value function u_1. Then, considering (for instance) the cost criterion with criteria 3, 4 and 5 in turn, one obtains a re-coding of each g_i into a single-attribute value function u_i.

The trade-off between u_1 and u_2 is easily determined through solving the following equation that just expresses the initial indifference in the standard sequence $(16\,500, 29.5) \sim (17\,500, 29.2)$

$$k_1 u_1(16\,500) + k_2 u_2(29.5) = k_1 u_1(17\,500) + k_2 u_2(29.2)$$

from which we get

$$\frac{k_2}{k_1} = \frac{u_1(16\,500) - u_1(17\,500)}{u_2(29.2) - u_2(29.5)}.$$

If we set k_1 to 1, this formula yields k_2 and the trade-offs k_3, k_4 and k_5 are obtained similarly. Notice that the re-coding process of the original evaluations into value functions results in a formulation in which all criteria have to be maximised (in value).

The above procedure, although rather intuitive and systematic is also quite complex; the questions are far from easy to answer; starting from one reference point or another (worst point instead of central point) may result in variations in the assessments. There are however many possibilities for checking for inconsistencies. Assume for instance that a single-attribute value function has been assessed by means of a standard sequence that links its scale to the cost criterion; one may validate this assessment by building a standard sequence that links its scale to another criterion and compare the two assessments of the same value function obtained in this way; hopefully they will be consistent; otherwise some sort of retroaction is required.

Note finally that such methods may not be used when the scale on which the assessments are made only has a finite number of degrees instead of being the set of real numbers; at least numerous and densely spaced degrees are needed.

Methods relying on numerical judgements

In another line of methods, simplicity and direct intuition are more praised than scrupulous satisfaction of theoretical requirements, although the theory is not ignored. An example is SMART ("Simple Multi-Attribute Rating Technique"), developed by W. Edwards, which is more a collection of methods than a single one. We just outline here a variant referring to von Winterfeldt and Edwards (1986), pp. 278 sq., for more details. In order to re-code, say, the evaluations for the acceleration criterion, one initially fixes two "anchor" points that may be the extreme values of the evaluations on the set of acceptable cars, here 28 and 31 seconds. On the value scale, the anchor points are associated to the endpoints of a conventional interval of values, for instance 31 to 0 and 28 to 100. Since 29 seconds seems to be the value under which Thierry considers that a car becomes definitely attractive from the acceleration viewpoint, it should be assigned to the interval $[28, 29]$ a range of values larger than $\frac{1}{3}$, its size (in relative terms) in the original scale. Thierry could for instance assign 29 seconds to 50 on the value scale. Then 28.5 and 30 could be located respectively in 70 and 10, yielding the initial sketch of a value function shown on Figure 6.6(a), (with linear interpolation between the specified values. This picture can be further improved by asking Thierry to see whether the relative spacings of the locations correctly reflect the strength of his preferences. Thierry might say that almost the same gain in value (40) from 30 seconds to 29 as from 29 to 28 (gain of 50) is unfair and he could consequently propose to lower to 40 the value associated with 29 seconds; he also lowers to 65 the value of 28.5 seconds. Suppose he is then satisfied with all other differences of values; the final version is drawn in Figure 6.6(b). A similar work has to be carried over for all criteria and the weights must be assessed.

The weights are usually derived through direct numerical judgements of relative attribute importance. Thierry would be asked to rank-order the attributes; an

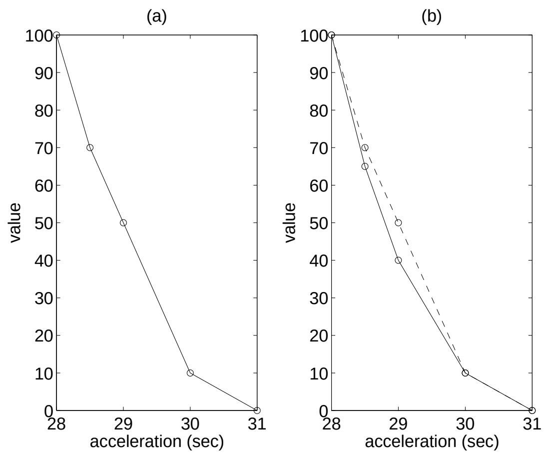


Figure 6.6: Value function for acceleration criterion: (a) initial sketch; (b) final, with initial sketch in dotted line

"importance" of 10 could be arbitrarily assigned to the least important criterion and the importance of each other criterion should be assessed in relation to the least important one, directly as an estimation of the ratio of weights. This approach in terms of "importance" can be and has been criticised. In assessing the relative weights no reference is made to the underlying scales. This is not appropriate since weights are trade-offs between units on the various value scales and must vary with the scaling.

For instance, on the acceleration value scale that is normalised in the 0-100 range, the meaning of one unit varies depending on the range of original evaluations (acceleration measured in seconds) that are represented between value 0 and value 100 of the value scale. If we had considered that the acceleration evaluations of admissible cars range from 27 to 32 seconds, instead of from 28 to 31, we would have constructed a value function u'_2 with $u'_2(32) = 0$ and $u'_2(27) = 100$; a difference of one unit of value on the scale u_2 illustrated in Figure 6.6 corresponds to a (less-than-unit) difference of $\frac{u'_2(28)-u'_2(31)}{100}$ on the scale u'_2. The weight attached to that criterion must vary in inverse proportion to the previous factor when passing from u_2 to u'_2. It is unlikely that a decision-maker would take the range of evaluations into account when asked to assess weights in terms of relative "importance" of criteria, a formulation that seems independent of the scalings of the criteria. A way of avoiding these difficulties is to give up the notion of importance that seems misleading in this context and to use a technique called *swing-weighting*; the decision-maker is asked to compare alternatives that "swing" between the worst and the best level for each attribute in terms of their contribution to the overall value. The argument of simplicity in favour of SMART is then lost since the questions to be answered are similar, both in difficulty and in spirit, to those raised in the approach based on indifference judgements.

6.3.2 AHP and Saaty's eigenvalue method

The eigenvalue method for assessing attribute weights and single-attribute value functions is part of a general methodology called "Analytic Hierarchy Process"; it consists in structuring the decision problem in a hierarchical manner (as it is also advocated for building value functions, for instance in Keeney and Raiffa (1976)), constructing numerical evaluations associated with all levels of the hierarchy and aggregating them in a specific fashion, formally a weighted sum of single-attribute value functions (see Saaty (1980), Harker and Vargas (1987)).

In our case, the top level of the hierarchy is Thierry's goal of finding the best car according to his particular views. The second level consists in the 5 criteria into which his global goal can be decomposed. The last level can be described as the list of potential cars. Thus the hierarchical tree is composed of 1 first level node, 5 second level nodes and 5 times 14 third level nodes also called leaves. What we have to determine is the "strength" or *priority* of each element of a level in relation to their importance for an element in the next level.

The assessment of the nodes may start (as is usually done) from the bottom nodes; all nodes linked to the same parent node are compared pairwise; in our case this amounts to comparing all cars from the point of view of a criterion and repeating this for all criteria. The same is then done for all criteria in relation to the top node; the influence of all criteria on the global goal are also compared pairwise. At each level, the pairwise comparison of the nodes in relation to the parent node is done by means of a particular method that allows, to some extent, to detect and correct inconsistencies. For each pair of nodes a, b, the decision-maker is asked to assess the "priority" of a as compared to the "priority" of b. The questions are expressed in terms of "importance" or "preference" or "likelihood" according to the context. It is asked for instance how much alternative a is preferred to alternative b from a certain point of view. The answers may be formulated either on a verbal or a numerical scale. The levels of the verbal scale correspond to numbers and are dealt with as such in the computations. The conversion of verbal levels into numerical levels is described in Table 6.6. There are five main levels on the verbal scale, but 4 intermediary levels that correspond to numerical codings $2, 4, 6, 8$ can also be used. For instance, the level "Moderate" corresponds to an alternative that is preferred 3 times more than another or a criterion that is 3 times more important than another. Such an interpretation of the verbal levels has very strong implications; it means that preference, importance and likelihood are considered as perceived on a ratio scale (much like sound intensity). This is indeed Saaty's basic assumption; what the decision-maker expresses as a level on the scale is postulated to be the ratio of values associated to the alternatives or the criteria. In other words, a number $f(a)$ is assumed to be attached to all a; when comparing a to b, the decision-maker is assumed to give an approximation of the ratio $\frac{f(a)}{f(b)}$. Since verbal levels are automatically translated into numbers in Saaty's method, we shall concentrate on assessing directly on the numerical scale.

Let $\alpha(a, b)$ denote the level of preference (or of relative importance) of a over b expressed by the decision-maker; the results of the pairwise comparisons may thus be encoded in a square matrix α. If Saaty's hypotheses are correct, there should

Verbal	Equal	Moderate	Strong	Very strong	Extreme
Numeric	1	3	5	7	9

Table 6.6: Conversion of verbal levels into numbers in Saaty's pairwise comparison method; e.g. "Moderate" means "3 times more preferred"

be some sort of consistency between elements of α, namely, for all a, b, c,

$$\alpha(a,c) \approx \alpha(a,b) \times \alpha(b,c) \tag{6.9}$$

and in particular,

$$\alpha(a,b) \approx \frac{1}{\alpha(b,a)} \tag{6.10}$$

In view of the latter relation, only one half (roughly) of the matrix has to be elicited, which amounts to answering $\frac{n(n-1)}{2}$ questions.

Relation (6.9) implies that all columns of matrix α should be approximately proportional to f. The pairwise comparisons enable to

1. detect departure from the basic hypothesis in case the columns of α are too far from proportional;
2. correct errors made in the estimation of the ratios; some sort of averaging of the columns is performed yielding an estimation of f.

A test based on statistical considerations allows the user to determine whether the assessments in the pairwise comparison matrix show sufficient agreement with the hypothesis that they are approximations of $\frac{f(a)}{f(b)}$, for an unknown f. If the test conclusion is negative, it is recommended either to revise the assessments or to choose another approach more suitable for the type of data.

If one wants to apply AHP in a multiple criteria decision problem, pairwise comparisons of the alternatives must be performed for each criterion; criteria must also be compared in a pairwise manner to model their importance. This process results in functions u_i that evaluate the alternatives on each criterion i and in coefficients of importance k_i. Each alternative a is then assigned an overall value $v(a)$ computed as

$$v(a) = \sum_{i=1}^{n} k_i u_i(a) \tag{6.11}$$

and the alternatives can be ranked according to the values of v.

Applying AHP to the case

Since Thierry did not apply AHP to his analysis of the case, we have answered the questions on pairwise comparisons on the basis of the information contained in his report. For instance, when comparing cars on the cost criterion, more weight will be put on a particular cost difference, say 1 000 €, when located in the range from 17 500 € to 21 500 € than when lying between 13 500 € and 17 500 €. This corresponds to the fact that Thierry said he is rather insensitive to cost differences up to about 17 500 €, which is the amount of money he had budgeted for his car. For the sake of concision, we have restricted our comparisons to a subset of cars, namely the top four cars plus the Renault 19, Mazda 323 and Toyota Corolla.

A major issue in the assessment of pairwise comparisons, for instance of alternatives in relation to a criterion, is to determine how many times a is preferred to b on criterion i from looking at the evaluations $g_i(a)$ and $g_i(b)$. Of course the (ratio) scale of preference on i is not in general the scale of the evaluations g_i. For example, Car 11 costs approximately 17 500 € and Car 12 costs about 16 000 €. The ratio of these costs, $\frac{17\,500}{16\,000}$, is equal to 1.09375 but this does not necessarily mean that Car 12 is preferred 1.09375 times more than Car 11 on the cost criterion; this is because the cost evaluation does not measure the preferences directly. Indeed, a transformation (re-scaling) is usually needed to go from evaluations to preferences; for the cost, according to Thierry himself, the transformation is not linear since equal ratios corresponding to costs located either below or above 17 500 € do not correspond to equal ratios of preference. But even in linear parts, the question is not easily answered. A decision-maker might very well say that Car 12 is 1.5 times more preferred than Car 11 for the cost criterion; or he could say 2 times or 4 times. All depends on what the decision-maker would consider as the minimum possible cost; for instance (supposing that the transformation of cost into preference is linear), if Car 12 is declared to be 1.5 times more preferred to Car 11, the zero of the cost scale x would be such that

$$\frac{17\,500 - x}{16\,000 - x} = 1.5\ ,$$

i.e. $x = 14\,500$ €. The problem is even more crucial for transforming scales such as those on which braking or road-holding are evaluated. For instance, how many times is Car 3 preferred to Car 10 with respect to the braking criterion? In other words, how many times is 2.66 better than (preferred to) 2.33?

Similar questions arise for the comparison of importance of criteria. We discuss the determination of the "weights" k_i of the criteria in formula 6.11 below. For computing those weights, the relative importance of each criterion with respect to all others must be assessed. Our assessments are shown in Table 6.7. We made them directly in numerical terms taking into account a set of weights that Thierry considered as reflecting his preferences; those weights have been obtained using the Prefcalc software and a method that is discussed in the next session. By default, the blanks on the diagonal should be interpreted as 1's; the blanks below the diagonal are supposed to be 1 over the corresponding value above the diagonal, according to equation 6.10.

Once the matrix in Table 6.7 has been filled, several algorithms can be proposed to compute the "priority" of each criterion with respect to the goal symbolised by the top node of the hierarchy (under the hypothesis that the elements of the assessment matrix are approximations of the ratios of those priorities). The most famous algorithm, which was initially proposed by Saaty, consists in computing the eigenvector of the matrix corresponding to the largest eigenvalue (see Harker

Relative importance	Cost	Accel	Pick-up	Brakes	Road-h
Cost		1.5	2	3	3
Acceleration			1.5	2	2
Pick-up				1.5	1.5
Brakes					1
Road-holding					

Table 6.7: Assessment of the comparison of importance for all pairs of criteria. For instance, the number 2 at the intersection of 1st row and 3rd column means that "Cost" is considered twice as important as "Pick-up'

and Vargas (1987), for an interpretation of the "eigenvector method" as a way of "averaging ratios along paths"). Since eigenvectors are determined up to a multiplicative factor, the vector of priorities is the normalised eigenvector whose components sum up to unity; the special structure of the matrix (*reciprocal* matrix) guarantees that all priorities will be positive. Alternative methods for correcting inconsistencies have been elaborated; most of them are based on some sort of a least squares criterion or on computing averages (see e.g. Barzilai, Cook and Golany (1987) who argue in favour of a geometric mean). Applying the eigenvector method to the matrix in Table 6.7, one obtains the following values that reflect the importance of the criteria:

$$(.352, .241, .172, .117, .117)$$

.

Note that only the lowest degrees of the 1 to 9 scale have been used in Table 6.7. This means that the weights are not perceived as very contrasted; in order to get the sort of gradation of the weights as above (the ratio of the highest to the lowest value is about 3), some comparisons have been assessed by non-integer degrees, which normally are not available on the verbal counterpart of the 1 to 9 scale described in Table 6.6. When the assessments are made through this verbal scale, approximations should be made, for instance by saying that cost and acceleration are equally important and substituting 1.5 by 1. Note that the labelling of the degrees on the verbal scale may be misleading; one would quite naturally qualify the degree to which "Cost" is more important than "Acceleration" as "Moderate" until it is fully realised that "Moderate" means "three times as important"; using the intermediary level between "Equal" and "Moderate" would still mean "twice as important".

It should be emphasised that the "eigenvalue method" is not linear. What would have changed if we had scaled the importance differently, for instance assessing the comparisons of importance by degrees twice as large as those in Table 6.7 (except for 1's that remain constant)? Would the coefficients of importance have been twice as large? Not at all! The resulting weights would have been much more contrasted, namely:

$$(.489, .254, .137, .060, .060)\ .$$

Name of car	Nr	7	11	3	12	10	4	6
Honda Civic	7	1.0	1.0	2.0	4.0	4.0	5.0	5.0
Peugeot 309/16V	11	1.0	1.0	2.0	3.0	4.0	4.0	4.0
Nissan Sunny	3	0.50	0.50	1.0	1.50	2.0	3.0	3.0
Peugeot 309	12	0.25	0.33	0.67	1.0	1.0	2.0	2.0
Renault 19	10	0.25	0.25	0.5	1.0	1.0	1.0	1.5
Mazda 323	4	0.2	0.25	0.33	0.5	1.0	1.0	1.0
Toyota Corolla	6	0.2	0.25	0.33	0.5	0.67	1.0	1.0

Table 6.8: Pairwise comparisons of preferences of 7 cars on the acceleration criterion

Using the latter set of weights instead of the former would substantially change the values attached to the alternatives through formula 6.11 and might even alter their ordering. So, contrary to the determination of the trade-offs in an additive value model (which may be re-scaled through multiplying them by a positive number, without altering the way in which alternatives are ordered by the multi-attribute value function), there is no degree of freedom in the assessment of the ratios in AHP; in other words, these assessments are made on an *absolute* scale.

As a further example, we now apply the method to determine the evaluation of the alternatives in terms of preference on the "Acceleration" criterion. Suppose the pairwise comparison matrix has been filled as shown in Table 6.8, in a way that seems consistent with what we know of Thierry's preferences. Applying the eigenvalue method yields the following "priorities" attached to each of the cars in relation to acceleration:

$$(.2987, .2694, .1507, .0934, .0745, .0584, .0548).$$

A picture of the resulting re-scaling of that criterion is provided in Figure 6.7; the solid line is a linear interpolation of the priorities in the eigenvector. A re-scaling of the same criterion had been obtained through the construction of a standard sequence (see Figure 6.5). Comparing these scales is not straightforward. Notice that the origin is arbitrary in the single-attribute value model; one may add any constant number to the values without changing the ranking of the alternatives (a term equal to the constant number times the trade-off associated to the attribute would just be added to the multi-attribute value function); since trade-offs depend on the scaling of their corresponding single-attribute value function, changing the unit on the vertical axis amounts to multiplying u_i by a positive number; the corresponding trade-off must then be divided by the same number. In the multi-attribute value model, the scaling of the single-attribute value function is related to the value of the trade-off; transformation of the former must be compensated for by transforming the latter. In AHP since the assessments of all nodes are made independently, no transformation is allowed. In order to compare the two figures, one may transform the value function of Figure 6.5 so it coincides with AHP priority on the extreme values of the acceleration half range, i.e. 28 and 29.5. Figure 6.7 shows the transformed single-attribute value function superimposed

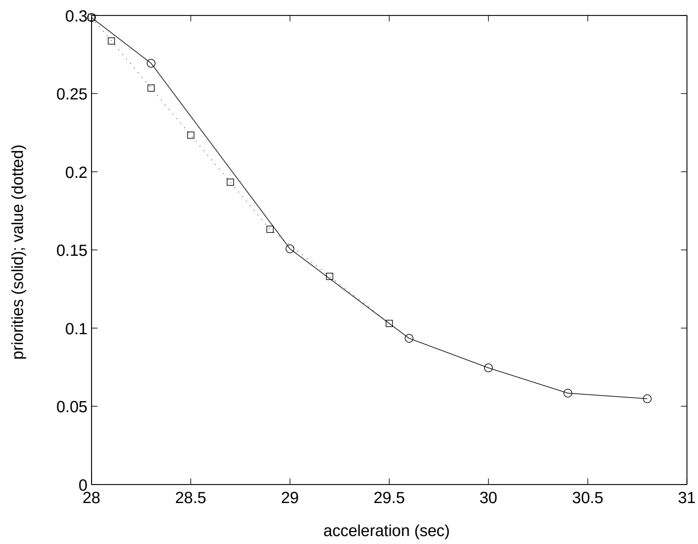


Figure 6.7: Priorities relatively to acceleration as obtained through the eigenvector method are represented by the solid line; the linearly transformed single-attribute values of Figure 6.5 are represented by the dotted line on the range from 28 to 29.5 seconds

(dotted line) on the graph of the priorities. There seems to be a good fit of the two curves but this is only an example from which no general conclusion can be drawn.

Comments on AHP

Although the models for describing the overall preferences of the decision-maker are identical in multi-attribute value theory and in AHP, this does not mean that applying the respective methodologies of these theories normally yields the same overall evaluation of the alternatives. There are striking differences between the two approaches from the methodological point of view. The ambition of AHP is to help construct evaluations of the alternatives for each viewpoint (in terms of preferences) and of the viewpoints with regard to the overall goal (in terms of importance); these evaluations are claimed to belong to a ratio scale, i.e. to be determined up to a positive multiplicative constant. Since the eigenvalue method yields a particular determination of this constant and this determination is not taken into account when assessing the relative importance of the various criteria, the evaluations in terms of preference must be considered as if they were made on an absolute scale, which has been repeatedly criticised in the literature (see for instance Belton (1986) and Dyer (1990)). This weakness (that can also be blamed on direct rating techniques, as mentioned above) could be corrected by asking the decision-maker about the relative importance of the viewpoints in terms of passing from the least preferred value to the most preferred value on criterion i compared to a similar change on criterion j (Dyer 1990). Taking this suggestion into account would however go against one of the basic principles of Saaty's methodology, i.e. the assumption that the assessments at all levels of the hierarchy can be made along the same procedure and independently of the other levels. That is probably why the original method, although seriously attacked, has remained unchanged.

AHP has been criticised in the literature in several other respects. Besides the fact already mentioned that it may be difficult to reliably assess comparisons of preferences or of importance on the standard scale described in Table 6.6, there is an issue about AHP that has been discussed quite a lot, namely the possibility of *rank reversal*. Suppose alternative x is removed from the current set and nothing is changed to the pairwise assessments of the remaining alternatives; it may happen that an alternative, say, a among the remaining ones could now be ranked below an alternative b whilst it was ahead of b in the initial situation. This phenomenon was discussed in Belton and Gear (1983) and Dyer (1990) (see also Harker and Vargas (1987) for a defense of AHP).

6.3.3 An indirect method for assessing single-attribute value functions and trade-offs

Various methods have been conceived in order to avoid direct elicitation of a multi-attribute value function. A class of such methods consists in postulating an additive value model (as described in formulae 6.7 and 6.8) and inferring all together the shapes of all single-attribute value functions and the values of all the trade-offs from declared global preferences on a subset of well-known alternatives. The idea is thus to infer a general preference model from partial holistic information about the decision-maker's preferences.

Thierry used a method of *disaggregation* of preferences described in Jacquet-Lagrèze and Siskos (1982); it is implemented in a software called *Prefcalc*, which computes piece-wise linear single-attribute value functions and is based on linear programming (see also Jacquet-Lagrèze (1990), Vincke (1992)). More precisely, the software helps to build a function

$$u(a) = \sum_{i=1}^{n} u_i(g_i(a))$$

such that $a \succsim b \iff u(a) \geq u(b)$. Without loss of generality, the lowest (resp. highest) value of u is conventionally set to 0 (resp. 1); 0 (resp. 1) is the value of an (fictitious) alternative whose assessment on each criterion would be to the worst (resp. best) evaluation attained for the criterion on the current set of alternatives. This fictitious alternative is sometimes called the *anti-ideal* (resp. *ideal*) point. In our example, the "anti-ideal" car, costs 21 334 €, needs 30.8 seconds to cover 1 km starting from rest and 41.6 seconds, starting in fifth gear at 40km/h; its performance regarding brakes and road-holding are respectively 1.33 and 1.25. The "ideal car" on the opposite side of the range, costs 13 841 €, needs 28 seconds to cover 1km starting from rest and 34.7 seconds, starting in fifth gear at 40km/h; its performance regarding brakes and road-holding are respectively 2.66 and 3.25.

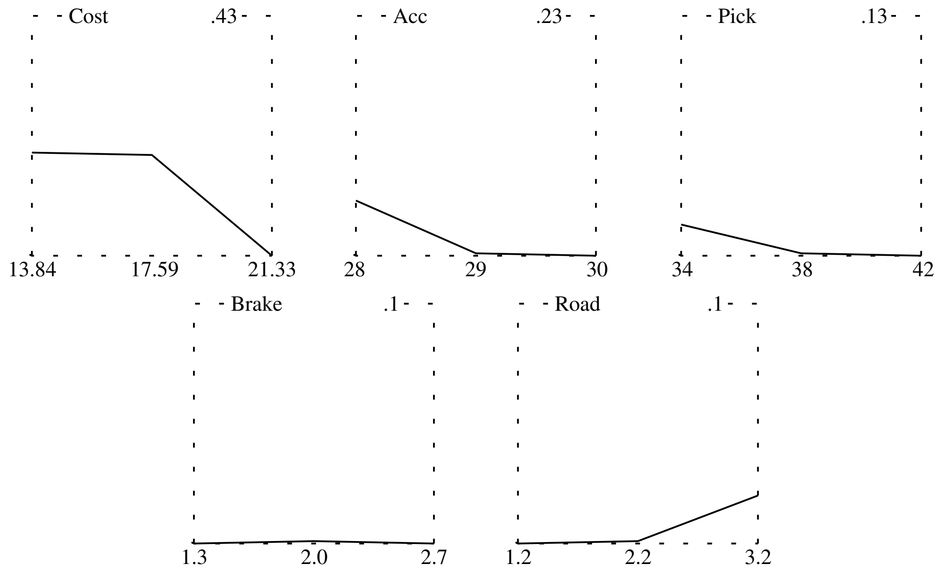


Figure 6.8: Single-attribute value functions computed by means of *Prefcalc* in the "Choosing a car" problem; the value of the trade-off is written in the right upper corner of each box

The shape of the single-attribute value function for the cost criterion for instance is modelled as follows. The user fixes the number of linear pieces; suppose that you decide to set it to 2 (which is a parsimonious option and the default value proposed in *Prefcalc*); the single-attribute value function of the cost could for instance be represented as in Figure 6.8. Note that the maximal value of the utility (reached for a cost of 13 841 €) is scaled in such a way that it corresponds to the value of the trade-off associated with the cost criterion, i.e. .43 in the example shown in Figure 6.8. Note also that with two linear pieces, one for each half of the cost range, the single-attribute value function is completely determined by two numbers, i.e. the utility value at mid-range and the maximal utility. Those values, say $u_{1,1}, u_{1,2}$ are variables of the linear program that *Prefcalc* writes and solves. The pieces of information on which the formulation of the linear program relies are obtained from the user. The user is asked to select a few alternatives that he is familiar with and feels able to rank-order according with his overall preferences. The ordering of these alternatives, which include the fictitious ideal and anti-ideal ones, induces the corresponding order on their overall value and hence, generates constraints of the linear program. *Prefcalc* then tries to find levels $u_{i,1}, u_{i,2}$ for each criterion i, which will make the additive value function compatible with the declared information. If the program is not contradictory, i.e. if an additive value function (with 2-piece piece-wise linear single-attribute value functions) proves compatible with the preferences, the system tries to find a solution among all feasible solutions, that maximises the discrimination between the selected alternatives. If no feasible solution can be found, the system proposes to increase the number of variables of the model, for instance by using a higher number of linear pieces in the description of the single-attribute value functions.

This method could be described as a *learning process*; the system fits the parameters of the model on the basis of partial information about the user's preferences; the set of alternatives on which the user declares his global preferences may be viewed as a learning set. For more details on the method, the reader is referred to Vincke (1992), Jacquet-Lagrèze and Siskos (1982).

In his ex post study Thierry selects five cars, besides the ideal and anti-ideal ones and ranks them in the following order:

1. Peugeot 309 GTI 16 (Car 11)
2. Nissan Sunny (Car 3)
3. Mitsubishi Galant (Car 13)
4. Ford Escort (Car 9)
5. Renault 21 (Car 14)

This ranking is compatible with an additive value function. Such a compatible value function is described in Figure 6.8.

Thierry examines this result and makes the following comments. He agrees with many features of the fitted single-attribute value functions and in particular with,

1. the lack of sensitivity in the price in the range from 13 841 € to 17 576 € (he was *a priori* estimating his budget at about 17 500 €);
2. the high importance (weight = .23) given to approaching 28 seconds on the "acceleration" criterion (above 29 seconds, the car is useless since a difference of 1 second in acceleration results in the faster car being two car lengths ahead of the slower one at the end of the test; Thierry declares this criterion to be the second most important after cost (weight = .43);
3. the importance (weight = .13) of getting as close as possible to 34 seconds in the acceleration test starting from 40 km/h (above 38 seconds he agrees that the car loses all attractiveness; the car is not only used in competition; it must be pleasant in everyday use and hence, the third criterion has a certain importance although it is of less importance than the second one);
4. the modelling of the road-holding criterion.

However, Thierry disagrees with the modelling of the braking criterion, which he considers equally important as road-holding. He believes that the relative importance of the fourth and fifth criteria should be revised. Thierry then looks at the ranking of the cars according to the computed value function. The ranking as well as the multi-attribute value assigned to each car are given in Table 6.9.

Rank		Cars	Value
1	*	Peugeot 309/16 (Car 11)	0.84
2	*	Nissan Sunny (Car 3)	0.68
3		Renault 19 (Car 10)	0.66
4		Peugeot 309 (Car 12)	0.65
5		Honda Civic (Car 7)	0.61
6		Fiat Tipo (Car 1)	0.54
7		Opel Astra (Car 8)	0.54
8		Mitsubishi Colt (Car 5)	0.53
9		Mazda 323 (Car 4)	0.52
10		Toyota Corolla (Car 6)	0.50
11		Alfa 33 (Car 2)	0.49
12	*	Mitsubishi Galant (Car 13)	0.48
13	*	Ford Escort (Car 9)	0.32
14	*	R 21 (Car 14)	0.16

Table 6.9: Ranking obtained using *Prefcalc*. The cars ranked by Thierry are those marked with a *

Thierry feels that Car 10 (Renault 19) is ranked too high while Car 7 (Honda Civic) should be in a better position.

In view of these observations, Thierry modifies the single-attribute value functions for criteria 4 and 5. For the braking criterion, the utility (0.01) associated with 2 remains unchanged while the utility of the level 2.7 is raised to 0.1 instead of 0.01. The road-holding criterion is also modified; the value (0.2) associated with the level 3.2 is lowered to 0.1 (see Figure 6.9). Note that *Prefcalc* normalises the value function in order that the ideal alternative is always assigned the value 1; of course due to the numbers display format with two decimal positions, the sum of the maximal values of the single-attribute value functions may be only approximately equal to 1. Running *Prefcalc* with the altered value functions returns the ranking in table 6.10 and the revised multi-attribute value after each car name.

After he sees the modified ranking yielded by *Prefcalc*, Thierry feels that the new ranking is fully satisfactory. He observes that if he had used *Prefcalc* a few years earlier, he would have made the same choice as he actually did; he considers this as a good point as far as *Prefcalc* is concerned. He finally makes the following comments: "Using *Prefcalc* has enhanced my understanding of both the data and my own preferences; in particular I am more conscious of the relative importance I give to the various criteria".

Comments on the method

First let us emphasise an important psychological aspect of the empirical validation of a method or a tool, which is common in human practice: the fact that previous intuition or previous more informal analyses are confirmed by using a tool, here *Prefcalc*, contributes to raising the level of confidence the user puts in the tool.

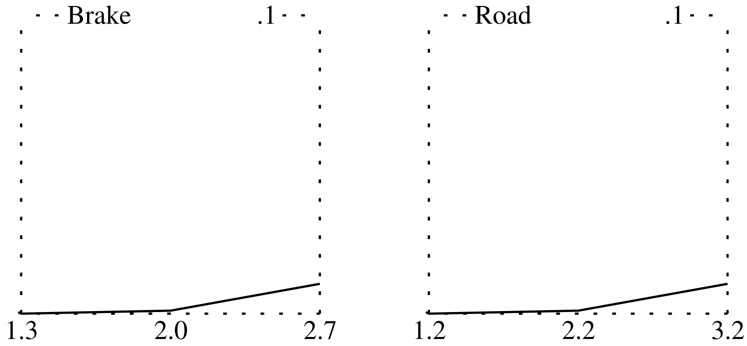


Figure 6.9: Modified single-attribute value functions for the braking and road-holding criteria

Rank		Cars	Value
1	*	Peugeot 309/16 (Car 11)	0.85
2	*	Nissan Sunny (Car 3)	0.75
3		Honda Civic (Car 7)	0.66
4		Peugeot 309 (Car 12)	0.65
5		Renault 19 (Car 10)	0.61
6		Opel Astra (Car 8)	0.55
7		Mitsubishi Colt (Car 5)	0.54
8		Mazda 323 (Car 4)	0.53
9		Fiat Tipo (Car 1)	0.51
10		Toyota Corolla (Car 6)	0.50
11	*	Mitsubishi Galant (Car 13)	0.48
12		Alfa 33 (Car 2)	0.47
13	*	Ford Escort (Car 9)	0.32
14	*	R 21 (Car 14)	0.16

Table 6.10: Modified ranking using *Prefcalc*. The cars ranked by Thierry are those marked with *

Observe that the user may well have a very vague understanding of the method itself; he simply validates the method by using it to reproduce results that he has confidence in. After such a successful empirical validation step he will be more prone to use the method in new situations that he does not master that well.

What are the drawbacks and traps of *Prefcalc*? Obviously *Prefcalc* can only be used in cases where the overall preference of the decision-maker can be represented by an additive multi-attribute value function (as described by Equation 6.8). In particular, this is not the case when preferences are not transitive or not complete (for arguments supporting the possible observation of non-transitive preferences, see the survey by Fishburn (1991)). There are some additional restrictions due to the fact that the shapes of the single-attribute value functions that can be modelled by *Prefcalc* are limited to piece-wise linear functions. This is hardly a restriction when dealing with a finite set of alternatives; by adapting the number of linear pieces one can obtain approximations of any continuous curve that can be as accurate as desired. When bounded to a small number of pieces, this may however be a more serious restriction.

Stability of ranking

The main problem raised by the use of such a tool is the indetermination of the estimated single-attribute value functions (including the estimation of the trade-offs). Usually, if the preferences declared on the set of well-known alternatives are compatible with an additive value model, there will be several value functions that can represent these preferences. *Prefcalc* chooses one such representation according to the principles outlined above, i.e. the most discriminating (in a sense). Other choices of a model albeit compatible with the declared preferences on the learning set, may lead to variations in the rankings of the remaining alternatives. Slight variations in the trade-off values can yield rank reversals. For instance, with all trade-offs within $\pm.02$ of their value in Figure 6.9, changes already occur. Passing from the set of trade-offs $(.43, .23, .13, .10, .10)$ to $(.45, .21, .11, .12, .10)$ results in exchanging the positions of Honda Civic and Peugeot 309, which are ranked 3rd and 4th respectively after the change. This rank reversal is obtained by putting slightly more emphasis on cost and slightly less on performance. Note that such a slight change in the trade-offs has an effect on the ranking of the top 4 cars, those on which Thierry focused after his preliminary analysis (see Table 6.3). It should thus be very clear that in practice, determining the trade-offs with sufficient accuracy could be both crucial and challenging. It is therefore of prime importance to carry out a lot of sensitivity analyses in order to identify which parts of the result remain reasonably stable.

Dependence on the learning set

In view of the fact that small variations of the trade-offs may even result in changes in the ranking of the top alternatives, one may question the influence of the selection of a learning set. In the case under examination, the top two alternatives were chosen to be in the learning set and hence, are constrained to appear in the correct order in the output of *Prefcalc*. What would have happened if the learning set had been different?

Let us take another subset of 5 cars and declare preferences that agree with the ranking validated by Thierry (Table 6.10). When substituting the top 2 cars (Peugeot 309/16V, Nissan Sunny) by Renault 19, Mitsubishi Colt, two cars in the middle segment of the ranking, the vector of trade-offs is (.53, .06, .08, .08, .25) and the top four in the new ranking are Renault 19 (1), Peugeot 309 (2), Peugeot 309/16V (3), and Nissan Sunny (4); Honda Civic is relegated to the 12th position. In the choice of the present learning set, stronger emphasis has been put on cost and safety (brakes and road-holding) and much less on performance (acceleration and pick up); three of the former top cars remain in the top four; Honda recedes due to its higher cost and its weakness on road-holding; Renault 19 is heading the race mainly due to excellent road-holding.

Further experiments have been performed, reintroducing in turn one of the 4 top cars and removing Renault 19. Clearly, the value of the trade-offs may depend drastically on the learning set. Some sort of preliminary analysis of the user's preferences can help to choose the learning set or understand the variations in the ranking and the trade-offs a posteriori. In the present case, one can be relatively satisfied with the results since the top 3 cars are usually well-ranked; the ranking of the Honda Civic is much more unstable and it is not difficult to understand why (weakness on road-holding and relatively high cost). The Renault 19 appears as an outsider due to excellent road-holding. Of course for the rest of the cars huge variations may appear in their ranking, but one is usually more interested in the top ranked alternatives.

From a general point of view, the option implemented in the mathematical programming model to reduce the indeterminacies (essentially, by choosing to maximise the contrast between the evaluations of the alternatives in the learning set) is not aimed at being as insensitive as possible with regard to the selection of a learning set. Other options could be experimentally investigated in order to see whether some could consistently yield more stable evaluations. It should be noted however that stability, which may be a desirable property in the perspective of uncovering an objective model of preferences measurement, is not necessarily a relevant requirement when the goal is to exploit partial available information. One may expect that the decision-maker will naturally choose alternatives that he considers as clearly distinct from one another as members of the learning set; the analyst might alternatively instruct the decision-maker to do so. In a learning process, where, typically, information is incomplete, it must be decided how to complement the available facts by some arbitrary default assumptions. The information should then be collected while taking the assumptions made into account; one may consider that in the case of *Prefcalc*, the analyst's instructions of selecting alternatives that are as contrasted as possible, is in good agreement with the implementation options.

6.3.4 Conclusion

This section has been devoted to the construction of a formal model that represents preferences on a numerical scale. Such a model can only be expected to exist when preferences satisfy rather demanding hypotheses; it thus relies on firm theoretical bases, which is undoubtedly part of the intellectual appeal of the method. There is at least one additional advantage to theoretically well-founded decision models; such models can be used to legitimate a decision to persons that have not been involved in the decision making process. Once the hypotheses of the model have been accepted or proved valid in a decision context and provided the process of elicitation of the various parameters of the model has been conducted correctly, the decision becomes transparent.

The additive multi-attribute value model is rewarding, when established and accepted by the stake-holders, since it is directly interpretable in terms of decision; the best decision is the one the model values most (provided the imprecisions in the establishment of the model and the uncertainties in the evaluation information allow to discriminate at least between the top alternatives). The counterpart of the clear-cut character of the conclusions that can be drawn from the model is that establishing the model requires a lot of information and of a very precise and particular type. This means that the model may be inadequate not only because the hypotheses could not be fulfilled but also because the respondents might feel unable to answer the questions or because their answers might not be reliable. Indirect methods based on exploiting partial information and extrapolating it (in a recursive validation process) may help when the information is not available in explicit form; it remains that the quality of the information is crucial and that a lot of it is needed. In conclusion, direct assessment of multi-attribute value functions is a narrow road between the practical problem of obtaining reliable answers to difficult questions and the risks involved in building a model on answers to simpler but ambiguous questions.

In the next section we shall explore a very different formal approach that may be less demanding with regard to the precision of the information, but also provides less conclusive outputs.

6.4 Outranking methods

6.4.1 Condorcet-like procedures in decision analysis

Is there any alternative way of dealing with multiple criteria evaluation in view of a decision to the one described above for building a one-dimensional synthetic evaluation on some sort of super-scale? To answer this question (positively), inspiration can be gained from the voting procedures discussed in Chapter 2 (see also Vansnick (1986)). Suppose that each voter expresses his preferences through a complete ranking of the candidates. With Borda's method, each candidate is assigned a rank for each of the voters (rank 1 if candidate is ranked first by a voter, rank 2 if he is ranked second, and so on); the Borda score of a candidate is the sum of the ranks assigned to him by the voters; the winner is the candidate with

Cars	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	5	3	1	2	2	3	3	2	3	2	2	2	2	3
2	2	5	2	4	2	3	2	3	3	1	1	1	4	3
3	4	4	5	4	4	4	4	4	4	3	2	3	5	4
4	3	1	1	5	1	3	1	2	1	2	1	1	4	2
5	3	3	1	5	5	3	2	2	2	3	1	1	5	2
6	2	2	1	2	2	5	2	2	2	2	1	1	3	2
7	3	3	1	4	4	4	5	3	4	3	2	2	4	4
8	3	2	1	4	4	4	3	5	3	2	0	2	4	3
9	2	3	1	4	4	3	1	2	5	2	1	2	4	3
10	4	4	2	3	2	3	2	3	3	5	3	2	4	3
11	4	4	3	4	4	4	4	5	4	3	5	4	4	5
12	4	4	2	4	4	4	4	4	3	4	3	5	5	4
13	3	2	0	2	1	2	1	2	1	1	1	0	5	1
14	2	3	1	3	3	3	1	3	3	2	0	1	4	5

Table 6.11: Number of criteria in favour of a when compared to b for all pairs of cars a, b in the "Choosing a car" problem

the smallest Borda score. This method can be seen as a method of construction of a synthetic evaluation of the alternatives in multiple criteria decision analysis, the points of view corresponding to the voters and the alternatives to the candidates; all criteria-voters have equal weight and coding by the rank number of the position of the candidate in a voter's preference looks like a form of evaluation.

Condorcet's method consists of a kind of tournament where all candidates compare in pairwise "contests". A candidate is declared to be preferred to another according to a majority rule, i.e. if more voters rank him before the latter than the converse. The result of such a procedure is a preference relation on the set of candidates that in general is neither transitive nor acyclic. A further step is thus needed in order to exploit this relation in view of the selection of one or several candidates or in view of ranking all the candidates. This idea can of course be transposed in the multiple criteria decision context. We do this below, using Thierry's case again for illustrative purpose; we show how the problems raised by a direct transposition rather naturally lead to elementary "outranking methods".

For each pair of cars a and b, we count the number of criteria according to which a is at least as good as b. This yields the matrix given in Table 6.11; the elements of the matrix are integers ranging from 0 to 5. Note that we might have alternatively decided to count the criteria for which a is better than b, not taking into account criteria for which a and b are tied.

What we could call the "Condorcet preference relation" is obtained by determining for each pair of alternatives a, b whether or not there is a (simple) majority of criteria for which a is at least as good as b. Since there are 5 criteria, the majority is reached as soon as at least 3 criteria favour alternative a when compared to b. The preference matrix is thus obtained by substituting 1 to any number larger or equal to 3 in Table 6.11 and 0 to any number smaller than 3 yielding the

Cars	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	1	1	0	0	0	1	1	0	1	0	0	0	0	1
2	0	1	0	1	0	1	0	1	1	0	0	0	1	1
3	1	1	1	1	1	1	1	1	1	1	0	1	1	1
4	1	0	0	1	0	1	0	0	0	0	0	0	1	0
5	1	1	0	1	1	1	0	0	0	1	0	0	1	0
6	0	0	0	0	0	1	0	0	0	0	0	0	1	0
7	1	1	0	1	1	1	1	1	1	1	0	0	1	1
8	1	0	0	1	1	1	1	1	1	0	0	0	1	1
9	0	1	0	1	1	1	0	0	1	0	0	0	1	1
10	1	1	0	1	0	1	1	1	1	1	1	0	1	1
11	1	1	1	1	1	1	1	1	1	1	1	1	1	1
12	1	1	0	1	1	1	1	1	1	1	0	1	1	1
13	1	0	0	0	0	0	0	0	0	0	0	0	1	0
14	0	1	0	1	1	1	0	1	1	0	0	0	1	1

Table 6.12: Condorcet Preference relation for the "Choosing a car"problem. A "1" at the intersection of the a row and the b column means that a is rated not lower than b on at least 3 criteria

relation described by the 0-1 matrix in Table 6.12. Note that a criterion counts both in favour of a and in favour of b only if a and b are tied on that criterion; the relation is reflexive since any alternative is at least as good as itself along all criteria.

Majority rule and cycles

It is not immediately apparent that this relation has cycles and even cycles that go through all alternatives; an instance of such a cycle is 1, 7, 10, 11, 3, 12, 5, 2, 14, 8, 9, 4, 6, 13, 1. Obviously it is not straightforward to suggest a good choice on the basis of such a relation since one can find 3 criteria (out of 5) saying that 1 is at least as good as 7, 3 (possibly different) criteria saying that 7 is at least as good as 10, ..., and finally 3 criteria saying that 13 is at least as good as 1. How can we possibly obtain something from this matrix in view of our goal of selecting the best car? A closer look at the preference relation reveals that some alternatives are preferred to most others while some to only a few ones; among the former are alternatives 11 (preferred to all), 3 (preferred to all but one), 12 (preferred to all but 2), 7 and 10 (preferred to all but 3). The same alternatives appear as seldom beaten: 3 and 11 (only once, excluding by themselves), 12 (twice), then come 10 (5 times) and 7 (6 times).

To make things appear more clearly, by avoiding cycles as much as possible, one might decide to impose more demanding levels of majority in the definition of a preference relation. We might require that an alternative be at least better than another on 4 criteria. The new preference relation is shown in Table 6.13.

All cycles in the previous relation disappeared. When ranking the alternatives

Cars	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	1	0	1	0	0	0	0	0	0	0	0	1	0
3	1	1	1	1	1	1	1	1	1	0	0	0	1	1
4	0	0	0	1	0	0	0	0	0	0	0	0	1	0
5	0	0	0	1	1	0	0	0	0	0	0	0	1	0
6	0	0	0	0	0	1	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	1	0	1	0	0	0	1	1
8	0	0	0	1	1	0	0	1	0	0	0	0	1	0
9	0	0	0	0	0	0	0	0	1	0	0	0	1	0
10	1	1	0	0	0	0	0	0	0	1	0	0	1	0
11	1	1	0	1	1	0	1	1	1	0	1	1	1	1
12	1	1	0	1	1	1	1	1	0	1	0	1	1	1
13	0	0	0	0	0	0	0	0	0	0	0	0	1	0
14	0	0	0	0	0	0	0	0	0	0	0	0	0	1

Table 6.13: Condorcet preference relation for the "Choosing a car" problem. A "1" at the intersection of the a row and the b column means that a is rated not lower than b on at least 4 criteria

by the number of those they beat (i.e. are at least as good on 4 criteria or more) one sees that 3, 11 and 12 come in the first position (they are preferred to 10 other cars), then there is a big gap after which come 7, 8 and 10 that beat only 3 other cars. Conversely, there are two non-beaten cars, 3 and 11, then come 10 and 12 (beaten by one car); 7 is beaten by 3 cars.

In the present case, we see that the simple approach that was used essentially makes the same cars emerge as the methods used so far. There are at least two radical differences between approaches based on the weighted sum and some more sophisticated way of assessing each alternative by a single number that synthesises all the criteria values. One is that all criteria have been considered equally important; it is possible however to take information on the relative importance of the criteria into account as will be seen in section 6.4.3.

The second difference is more in the nature of the type of approach; the most striking point is that the size of the differences in the evaluations of a and b for all criteria does not matter; only the signs of those differences do. In other words, had the available information been rankings of the cars with respect to each criterion (instead of numeric evaluations), the result of the "Condorcet" procedure would have been exactly the same. More precisely, suppose that all that we know (or that Thierry considers relevant in terms of preferences) about the cost criterion is the ordering of the cars according to the estimated cost, i.e.

$$\text{Car } 6 \succ_1 \text{Car } 5 \succ_1 \text{Car } 2 \succ_1 \text{Car } 4 \succ_1 \text{Car } 12 \succ_1$$
$$\text{Car } 10 \succ_1 \text{Car } 3 \succ_1 \text{Car } 13 \succ_1 \text{Car } 11 \succ_1 \text{Car } 8 \succ_1$$
$$\text{Car } 1 \succ_1 \text{Car } 7 \succ_1 \text{Car } 9 \succ_1 \text{Car } 14$$

where $\succ_1$ represents " is preferred to ... on Criterion 1 ", i.e. " is cheaper than ... ". Suppose that similar hypotheses are made for the other 4 criteria; if this were the case we would have obtained the same matrices as in Tables 6.12 and 6.13. Of course, neglecting the size of the differences for a criterion such as cost may appear as misusing the available information; there are at least two considerations that could mitigate this commonsense reaction:

- the assessments for the cars on the cost criterion are rather rough estimations of an expected cost (see section 6.1.1); in particular it is presumed that on average the lifetimes of all alternatives are equal; is it reasonable in those circumstances to rely on precise values of differences of these estimations to select the "best" alternative?
- estimations of cost, even reliable ones, are not necessarily related with preferences on the cost criterion in a simple way.

Such issues were discussed extensively in section 6.2.4. The whole analysis carried out there was aimed towards the construction of a multiple criteria value function, which implies making any difference in evaluations on a criterion equivalent to some uniquely defined difference for any other criterion. The many methods that can be used to build a value function by questioning a decision-maker about his preferences may well fail however; let us list a few reasons for the possible failure of these methods:

- time pressure may be so intense that there is not enough time available to engage in the lengthy elicitation process of a multiple criteria value function;
- it may be that the importance of the decision to be made does not justify such an effort;
- the decision-maker might not know how to answer the questions or might try to answer but prove inconsistent or might feel discomfort in being forced to give precise answers where things are vague to him;
- in case of group decision, the analyst may be unable to make the various decision-makers agree on the answers to be given to some of the questions raised in the elicitation process.

In such cases it may be inappropriate or inefficient to try building a value function and other approaches may be preferred. This appears perhaps better if we consider the more artificial scales associated with criteria 4 and 5 (see section 6.1.1 concerning the construction of these scales). Take, for instance, criterion 4 (Brakes). Does the difference between the levels 2.33 and 2.66 have a quantitative meaning? If it does, is this difference, in terms of preferences, more than, less than or equal to the difference between the levels 1.66 and 2? How much would you accept to pay (in terms of criterion 1) to raise the value for criterion 4 from 2.33 to 2.66 or from 1.33 to 2.33? Of course questions raised for eliciting value functions are more indirect but they still require a precise perception of the meaning of the levels on the scale of criterion 4 by the decision-maker. Such a perception can only be obtained by having experienced the braking behaviour of specific cars rated at the various levels of the scale, but such knowledge cannot be expected from a decision-maker (otherwise there would be no room on the marketplace for all the magazines that evaluate goods in order to help consumers spend their money while making the best choice). Also remember that braking performance has been described by the average of 3 indices evaluating aspects of the cars' braking behaviour; this does not favour a deep intuitive perception of what the levels on that scale may really mean. So, one has to admit that in many cases the definition of the levels on scales is quite far from precise in quantitative terms and it may be "hygienic" not to use the fallacious power of numbers. This is definitely the option chosen in the methods discussed in the present section. Not that these methods are purely ordinal; but differences between levels on a scale are carefully categorised, yet usually in a coarse-grained fashion, in order not to take into account differences that are only due to the irrelevant precision of numbers.

6.4.2 A simple outranking method

The Condorcet idea for a voting procedure has been transposed in decision analysis under the name of *outranking* methods. Such a transposition takes the peculiarities of the decision analysis context into account, in particular the fact that criteria may be perceived as unequally important; additional elements such as the notion of *discordance* have also been added. The principle of these methods is as follows. Each pair of alternatives is considered in turn independently of third part alternatives; when looking at alternatives a and b, it is claimed that a "outranks" b if there are enough arguments to decide that a is at least as good as b, while there is no essential reason to refute that statement (Roy (1974), cited by Vincke (1992), p. 58). Note that taking strong arguments against declaring a preference into account is typically what is called "discordance" and is original with respect to the simple Condorcet rule. Such an approach has been operationalised through various procedures and particularly the family of ELECTRE methods associated with the name of B. Roy. (For an overview of outranking methods, the reader is referred to the books by Vincke (1992) and Roy and Bouyssou (1993)). Below, we discuss an application of the simplest of these methods, ELECTRE I, to Thierry's case; ELECTRE I is a tool designed to be used in the context of a choice decision problem; it builds up a set of which the best alternative—according to the decision-maker's preferences—should be a member. Let us emphasise that this set cannot be described as the set of best alternatives, not even a set of good alternatives, but just a set that contains the "best" alternatives. We shall then show how the fundamental ideas of ELECTRE I can be sophisticated, in particular in view of helping to rank the alternatives. Our goal is not to make a survey of all outranking methods; we just want to present the basic ideas of such methods and illustrate some problems they may raise.

The lack of transitivity, acyclicity and completeness issues

As a preamble, it may be useful to emphasise the fact that outranking methods (and more generally methods based on pairwise comparisons) do not generally yield preferences that are transitive (not even acyclic). This point was already made in Chapter 2 about Condorcet's method. Since the hypotheses of Arrow's theorem can be re-formulated to be relevant in the framework of multiple criteria decision analysis (through the correspondence candidate-alternative, voter-criterion; see also Bouyssou (1992) and Perny (1992)), it is no wonder that methods based on comparisons of alternatives by pairs, independently of the other alternatives, will seldom directly yield a ranking of the alternatives. The pairs of alternatives that belong to the outranking relation are normally those between which the preference is established with a high degree of confidence; contradictions are reflected either in cycles (a outranks b that outranks c that ... that outranks a) or incomparabilities (neither a outranks b nor the opposite).

Let us emphasise that the lack of transitivity or of completeness, although raising operational problems, may be viewed not as a weakness but rather as faithfully reflecting preferences as they can be perceived at the end of the study. Defenders of the approach support the idea that forcing preferences to be expressed in the format of a complete ranking is in general too restrictive; there is experimental evidence that backs their viewpoint (Tversky (1969), Fishburn (1991)). Explicit recognition that some alternatives are incomparable may be an important piece of information for the decision-maker.

In addition, as repeatedly stressed in the writings of B. Roy, the outranking relation should be interpreted as what is clear-cut in the preferences of the decision-maker, something like the surest and most stable expression of a complex, vague and evolving object that is named, for simplicity, "the preferences of the decision-maker". In this approach very little hypotheses are made on preferences (like rationality hypotheses); one may even doubt that preferences pre-exist the process from which they emerge.

The analysis of a decision problem is conceived as an *informational* process, in which, carefully, prudently and interactively, models are built that reflect, to some extent, the way of thinking, the feelings and the values of a decision-maker; in this concept, the concern is not making a decision but helping a decision-maker to make up his mind, helping him to understand a decision problem while taking his own values into account in the modelling of the decision situation.

The approach could be called *constructive*; it has many features in common with a learning process; however, in contrast with most artificial intelligence practice, the model of preferences is built explicitly and formally; preferences are not simply described through rules extracted from partial information obtained on a learning set. For more about the constructive approach including comparisons with the classical *normative* and *descriptive* approaches (see Bell, Raiffa and Tversky (1988)), the reader is referred to Roy (1993).

Once the outranking relation has been constructed, the job of suggesting a decision is thus not straightforward. A phase of exploitation of the outranking relation is needed in order to provide the decision-maker with information more directly interpretable in terms of a decision. Such a two-stage process offers the advantage of good control on the transformation of the multi-dimensional information into a model of the decision-maker's preferences including a certain degree of inconsistency and incompleteness.

6.4.3 Using ELECTRE I on the case

We briefly review the principles of the ELECTRE I method. For each pair of alternatives a and b, the so-called *concordance* index is computed; it measures the *strength of the coalition* of criteria that support the idea that a is at least as good as b. The strength of a coalition is just the sum of the *weights* associated to the criteria that constitute the coalition. The notion of weights will be discussed below. If all criteria are equally important, the concordance index is proportional to the number of criteria in favour of a as compared to b as in the Condorcet-like method discussed above. The level from which a coalition is judged strong enough is determined by the so-called concordance threshold; in the Condorcet voting method, with the simple majority rule, this threshold is just half the number of criteria and in general one will choose a number above half the sum of the weights of all criteria. Another feature that contrasts ELECTRE with pure Condorcet but also with purely ordinal methods, is that some large differences in evaluation, when in disfavour of a, might be pinpointed as preventing a from outranking b. One therefore checks whether there is any criterion for which b is so much better than a that it would make it meaningless for a to be declared preferred overall to b; if this happens for at least one criterion one says that there is a *veto* to the preference of a over b. If the concordance index passes some threshold ("concordance threshold") and there is no veto of b against a, then a outranks b. Note that the outranking relation is not asymmetric in general; it may happen that a outranks b and that b outranks a.

This process yields a binary relation on the set of alternatives, which may have cycles and be incomplete (neither a outranks b nor the opposite). In order to propose a set of alternatives of particular interest to the decision-maker from which the best compromise alternative should emerge, one extracts the *kernel* of the graph of the outranking relation after having the cycles reduced; in other words, all alternatives in a cycle are considered to be equivalent; they are substituted by a unique representative node; in the resulting relation without cycles, the kernel is defined as a subset of alternatives that do not outrank one another and such that each alternative not in the kernel is outranked by at least one alternative in the kernel; in particular all non-outranked alternatives belong to the kernel. In a graph without cycles, a unique kernel always exists. It should be emphasised that all alternatives in the kernel are not necessarily good candidates for selection; an alternative incomparable to all others is always in the kernel; alternatives in the kernel may be beaten by alternatives not in the kernel. So, the kernel may be viewed as a set of alternatives on which the decision-maker's attention should be focused.

In order to apply the method to Thierry's case, we successively have to determine

- weights for the criteria
- a concordance threshold
- ordered pairs of evaluations that lead to a veto (and this for every criterion)

Evaluating coalitions of criteria

The concordance index $c(a, b)$, that measures the coalition of criteria along which a is at least as good as b may be computed by the formula

$$c(a,b) = \sum_{i:g_i(a)\geq g_i(b)} p_i \tag{6.12}$$

where the p_i's are normalised weights that reflect the relative importance of the criteria; $g_i(a)$ denotes, as usual, the evaluation of alternative a for criterion i (which is assumed to be maximised; if it were to be minimised, the weight p_i would be added when the converse inequality holds, i.e. $g_i(a) \leq g_i(b)$). So, as often as the evaluation of a passes or equals that of b on a criterion, its weight now enters into the weight of the coalition (additively) in favour of a. A criterion can count both for a against b and the opposite if and only if $g_i(a) = g_i(b)$.

In the context of outranking, the weights are not trade-offs; they are completely independent of the scales for the criteria. A practical consequence is that one may question the decision-maker in terms of relative importance of the criteria without reference to the scales on which the evaluations for the various viewpoints are expressed. This does not mean however that they are independent of the method and that one could use values given spontaneously by the decision-maker or through questioning in terms of "importance" without care, without reference to the evaluations as is done in Saaty's procedure. It is important to bear in mind how the weights will be used, in this case to measure the strength of coalitions in pairwise comparisons and decide on the preference only on the basis of the coalitions.

To be more specific and contrast the meaning of the weights from those used in weighted sums, let us first consider those suggested by Thierry in section 6.2.2, i.e. $(1, 2, 1, 0.5, 0.5)$. Note that these were not obtained through questioning on the relative importance of criteria but in the context of the weighted sum with Thierry bearing re-scaled evaluations in mind: the evaluations on each criterion had been divided by the maximal value $g_{i,max}$ attained for that criterion. Dividing the weights by their sum (= 5), yields the normalised weights $(.2, .4, .2, .1, .1)$. Using these weights in outranking methods would lead to an overwhelming predominance of criteria 2 (Acceleration) and 3 (Pick-up), which are also linked since they are facets of the cars performance. With such weights and a concordance threshold of at least .5 , it is impossible for a car to be outranked when it is better on criteria 2 and 3 even if all other criteria are in favour of an opponent. It was never Thierry's intention that once a car is better on criteria 2 and 3, there is no need for looking at the other criteria; the whole initial analysis shows on the contrary, that a fast and powerful car is useless, for instance, if it is bad on the braking or road-holding criterion. Such a feature of the preference structure could indeed be reflected through the use of vetoes, but only in a negative manner, i.e. by removing the outranking of a safe car by a powerful one, not by allowing a safe car to outrank a powerful one. Note that the above weights may nevertheless be appropriate for a weighted sum because in such a method, the weights are multiplied by the evaluations (or re-coded evaluations). To make it clearer, consider the following reformulation of the condition under which a is preferred to b in the weighted sum model (a similar formulation is straightforward in the additive value model)

$$a \succsim b \text{ iff } \sum_{i=1}^{n} k_i \times (g_i(a) - g_i(b)) \geq 0. \tag{6.13}$$

If a is slightly better than b on a point of view i, the influence of this fact in the comparison between a and b is reflected by the term $k_i \times (g_i(a) - g_i(b))$ which is presumably small. Hence, important criteria count for little in pairwise comparisons when the difference between the evaluations of the alternatives are small enough. On the contrary, in outranking methods, weights are not divided; when a is better than b on some criterion, the full weight of the criterion counts in favour of a, whether a is either slightly or by far better than b.

Since the weights in a weighted sum depend on the scaling of each criterion and there is no acknowledged standard scaling, it makes no sense in principle to use the weights initially provided by Thierry as coefficients measuring the importance of the criteria in an outranking method. If we nevertheless try to use them, we might consider the weights used with the normalised criteria of Table 6.4. We see that the importance of the "safety coalition" (Criteria 4 and 5) would be negligible (weight = .20), while the importance of the "performance coalition" (Criteria 2 and 3) would be overwhelming (weight = .60). There is another reasonable normalisation of the criteria that does not fix the zero of the scale but rather maps the smallest attained value $g_{i,min}$ onto 0 and the largest $g_{i,max}$ onto 1. Transforming the weights accordingly (i.e. multiplying them by the inverse of the range of the values for the corresponding criterion prior to the transformation) one would obtain $(.28, .14, .13, .20, .25)$ as a weight vector. With these values as coefficients of importance, the "safety coalition" (Criteria 4 and 5; weight = .45) becomes more important than the "performance coalition" (Criteria 2 and 3; weight = .27) that Thierry may consider unfair. As an additional conclusion, one may note that the values of the weights vary tremendously depending on the type of normalisation applied.

Now look at the weights (.35, .24, .17, .12, .12) obtained through Saaty's questioning procedure in terms of "importance" (see section 6.3.2). Using these weights for measuring strength of coalitions does not seem appropriate, since criteria 1 and 2's predominance is too strong (joint weight = $.35 + .24 = .59$).

Due to the all or nothing character of the weights in ELECTRE I, one is inclined to choose less contrasted weights than those examined above. Although there are procedures that have been proposed to elicit such weights (see Mousseau (1993), Roy and Bouyssou (1993)), we will just choose a set of weights in an intuitive manner; let us take weights proportional to $(10, 8, 6, 6, 6)$ as reflecting the relative importance of the criteria. At least the ordering of the values seems to be

Cars	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	1	.5	.17	.33	.33	.56	.61	.33	.61	.33	.33	.33	.33	.61
2	.49	1	.44	.83	.33	.56	.44	.61	.61	.28	.28	.28	.83	.61
3	.83	.73	1	.73	.73	.73	.78	.78	.83	.56	.44	.56	1	.78
4	.66	.17	.28	1	.17	.56	.28	.44	.28	.44	.28	.28	.78	.44
5	.66	.66	.28	1	1	.56	.44	.44	.44	.66	.28	.28	1	.44
6	.44	.44	.28	.44	.44	1	.44	.44	.44	.44	.28	.28	.61	.44
7	.56	.56	.22	.73	.73	.73	1	.56	.83	.56	.39	.39	.73	.83
8	.66	.39	.22	.73	.73	.73	.61	1	.66	.39	0	.39	.73	.66
9	.39	.56	.17	.73	.73	.56	.17	.33	1	.39	.17	.39	.73	.61
10	.83	.73	.44	.56	.33	.56	.61	.61	.61	1	.61	.33	.83	.61
11	.83	.73	.56	.73	.73	.73	.78	1	.83	.56	1	.73	.73	1
12	.83	.73	.44	.73	.73	.73	.78	.78	.61	.83	.61	1	1	.78
13	.66	.39	0	.39	.17	.39	.28	.44	.28	.17	.28	0	1	.28
14	.39	.56	.22	.56	.56	.56	.17	.56	.56	.39	0	.22	.73	1

Table 6.14: Concordance index (rounded to two decimals) for the "Choosing a car" problem

in agreement with what is known about Thierry's perceptions. Normalising the weight vector yields $(.27, .22, .17, .17, .17)$ after rounding in such a way that the normalised weights sum up to 1.00. The weights of the three groups of criteria are rather balanced; .27 for cost, .39 for performance and .34 for safety. The concordance matrix $c(a, b)$ computed with these weights is shown in Table 6.14.

Determining which coalitions are "strong enough"

At this stage we have to build the concordance relation, a binary relation obtained through deciding which coalitions in Table 6.14 are strong enough; this is done by selecting a *concordance threshold* above which we consider that they are. If we set the concordance threshold at .60, we obtain a concordance relation with a cycle passing through all alternatives but one, which is Car 3. This tells us something about coalitions that we did not know. Previous analysis with equal weights (see Section 6.4.1) showed that the relation in Table 6.12, obtained through looking at concordant coalitions involving at least three criteria, had a cycle passing through all alternatives; with the weights we have now chosen, the "lightest" coalition of three criteria involves criteria 3, 4 and 5 and weighs .51; then, in increasing order, we have three different coalitions weighing .56 (two of the criteria 3, 4, 5 with criterion 2), and three coalitions weighing .61 (two of the criteria 3, 4, 5 with criterion 1); finally there are three coalitions weighing .66 (one of the three criteria 3, 4, 5 together with criteria 1 and 2). Cutting the concordance index at .60 thus only keeps the 3-coalitions that contain criterion 1 with the coalitions involving at least 4 criteria.

The new thing that we can learn is the following: the relation obtained by looking at coalitions of at least 4 criteria plus coalitions of three that involve criterion 1 has a big cycle. When we cut above .62 there is no longer a cycle. The "lightest" 4-coalition weighs .73 and there is only one value of the concordance index between .61 and .73, namely .66. So cutting between .66 and .72 will yield the relation in Table 6.13, which we have already looked at; a poorer relation (i.e. with fewer arcs) is obtained when cutting above .73. In the sequel we will concentrate on two values of the concordance threshold, .60 and .65, that are on both sides of the borderline separating concordance relations with and without cycles; above these values, concordance relations tend to become increasingly poor; below, they are less and less discriminating.

In the above presentation the weights sum up to 1. Note that multiplying all the weights by a positive number would yield the same concordance relations provided the concordance threshold is multiplied by the same factor; the weights in ELECTRE I may be considered as being assessed on a ratio scale, i.e. up to a positive scaling factor.

Supporting choice or ranking

Before studying discordance and veto we show how a concordance relation, which is just an outranking relation without veto, can be used for supporting a choice or a ranking in a decision process. Introducing vetoes will just remove arcs from the concordance relation but the operations performed on the outranking relation during the exploitation phase are exactly those that are applied below to the concordance relation.

In view of supporting a choice process, the exploitation procedure of ELECTRE I firstly consists in *reducing the cycles*, which amounts to consider all alternatives in a cycle as equivalent. The kernel of the resulting acyclic relation is then searched for and it is suggested that the kernel contains all alternatives on which the attention of the decision-maker should be focused. Obviously, reducing the cycles involves some drawbacks. For example, cutting the concordance relation of Table 6.14 at .60 yields a concordance relation with cycles involving all alternatives but Car 3; there is no simple cycle passing once through all alternatives except Car 3; an example of (non-simple) cycle is 1, 7, 9, 5, 10, 11, 12, 2, 14, 13, 1 plus, starting from 12, 12, 8, 4, 1 and again, 12, 6, 1. Reducing the cycles of this concordance relation results in considering two classes of equivalent alternatives; one class is composed of the single Car 3 while the other class comprises all other alternatives. Beside the fact that this partition is not very discriminating it also considers as equivalent alternatives that are not in the same simple cycle. Moreover, the information on how the alternatives compare with respect to all others is completely lost; for instance Car 12, which beats almost all other alternatives in the cut at .60 of the concordance relation, would be considered as equivalent to Car 6 which beats almost no other car.

For illustrative purposes, we consider the cut at level .65 of the concordance index, which is the largest acyclic concordance relation that can be obtained; this relation is shown in Table 6.15. Its kernel is composed of cars 3, 10 and 11. Cars 3 and 11 are not outranked and car 10 is the only alternative that is not outranked either by car 3 or by car 11. This seems to be an interesting set in a choice process, in view of the analysis of the problem carried out so far.

Rankings of the alternatives may also be obtained from Table 6.15 in a rather simple manner. For instance, consider the alternatives either in decreasing order of the number of alternatives they beat in the concordance relation or in increasing order of the number of alternatives by which they are beaten in the concordance

Cars	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	1	0	1	0	0	0	0	0	0	0	0	1	0
3	1	1	1	1	1	1	1	1	1	0	0	0	1	1
4	1	0	0	1	0	0	0	0	0	0	0	0	1	0
5	1	1	0	1	1	0	0	0	0	1	0	0	1	0
6	0	0	0	0	0	1	0	0	0	0	0	0	0	0
7	0	0	0	1	1	1	1	0	1	0	0	0	1	1
8	1	0	0	1	1	1	0	1	1	0	0	0	1	1
9	0	0	0	1	1	0	0	0	1	0	0	0	1	0
10	1	1	0	0	0	0	0	0	0	1	0	0	1	0
11	1	1	0	1	1	1	1	1	1	0	1	1	1	1
12	1	1	0	1	1	1	1	1	0	1	0	1	1	1
13	1	0	0	0	0	0	0	0	0	0	0	0	1	0
14	0	0	0	0	0	0	0	0	0	0	0	0	1	1

Table 6.15: Concordance relation for the "Choosing a car" problem with weights .28, .22, .17, .17, .17 and concordance threshold .65

Class	1	2	3	4	5	6	7	8	9
A	11	3, 12	8	7	5	9, 10	2, 4	13, 14	1, 6
	(11)	(10)	(7)	(6)	(5)	(3)	(2)	(1)	(0)
B	3, 11	12	10	7, 8	9	2, 6, 14	5	1, 4	13
	(0)	(1)	(2)	(3)	(4)	(5)	(6)	(8)	11

Table 6.16: Rankings obtained from counting how many alternatives are beaten (ranking "A") or beat (ranking "B") each alternative in the concordance relation (threshold .65); the numbers between parentheses in the second row of ranking A (resp. ranking B) are the numbers of beaten (resp. beating) alternatives for each alternative of the same column in the first row

relation. This amounts to counting the 1's respectively in rows and columns of Table 6.15 and ranking the alternatives accordingly (we do not count the 1's on the diagonal since the coalition of criteria saying that an alternative is at least as good as itself always encompasses all criteria); the corresponding rankings are respectively labelled "A" and "B" in Table 6.16. We observe that the usual group of "good" alternatives form the top two classes of these rankings.

There are more sophisticated ways of obtaining rankings from outranking relations. ELECTRE II, which we do not describe here, was designed for fulfilling this goal. To some extent, it makes better use of the information contained in the concordance index, since the ranking is based on two cuts, one linked with a weak preference threshold, the other, with a strong preference threshold; for instance in our case, one could consider that the .60 cut corresponds to weak preference (or weak outranking) while the .65 cut corresponds to strong preference. In the above method, the information contained in other cutting levels has been totally ignored although the rankings obtained from them may not be identical. They may even differ significantly as can be seen when deriving a ranking from the .60 cut by using the method we applied to the .65 cut.

Thresholding

To this point, both in the Condorcet-like method and the basic ELECTRE I method (without veto), we treated the assessments of the alternatives as if they were ordinal data, i.e. we could have obtained exactly the same results (kernel or ranking) by working with the orders induced from the set of alternatives by their evaluations on the various criteria. Does this mean that outranking methods are purely ordinal? Not exactly! More sophisticated outranking methods exploit information that is richer than purely ordinal but not as demanding as cardinal. This is done through what we shall call "thresholding". Thresholding amounts to identifying intervals on the criteria scales, which represent the minimal difference evaluation above which a particular property holds. For instance, consider that the assessment of b on criterion i, $g_i(b)$, is given and criterion i is to be maximised; from which value $g_i(b) + t_i(g_i(b))$ onwards, will an alternative a be said to be preferred to b? Implicitly, we have considered previously that b was preferred to a on criterion i as soon as $g_i(b) \geq g_i(a)$, i.e. we have considered that $t_i(g_i(b)) = 0$. In view of imprecision in the assessments and since it is not clear for all criteria that there is a marked preference when the difference $|g_i(a) - g_i(b)|$ is small, one may be led to consider a non-null threshold to model preference. In our case, for instance, it is not likely that Thierry would really mark a preference between cars 3 and 10 on the Cost criterion since their estimated costs are within 10 € (see Table 6.2). Thresholding is all the more important that, as mentioned at the end of section 6.4.1, the size of the interval between the evaluations is not taken into account when deciding that a is overall preferred to b. Hence one should be prudent when deciding that a criterion is or is not an argument for saying that a is at least as good as b; therefore, it is reasonable to determine a threshold function t_i and say that criterion i is such an argument as soon as $g_i(a) \geq g_i(b) + t_i(g_i(b))$; since we examine reasons for saying that a is at least as good as b, not for saying that a is (strictly) better than b, the function t_i should be negatively valued.

Determining such a threshold function is not necessarily an easy task. One could ask the decision-maker to tell, ideally for each evaluation $g_i(a)$ of each alternative on each criterion, from which value onwards an evaluation should be considered at least as good as $g_i(a)$. Things may become simpler if the threshold may be considered constant or proportional to $g_i(a)$ (e.g. $t_i(g_i(a)) = .05 \times g_i(a)$). Note that constant thresholds could be used when a scale is "linear" in the sense that equal differences throughout a scale have the same meaning and consequences (see end of section 6.2.3); however this is not a necessary condition since some differences, but not all, need to be equivalent throughout the scale. In any case, Definition 6.12 of the concordance index is adapted in a straightforward manner as follows and the method for building an outranking relation remains unchanged:

$$c(a,b) = \sum_{i:g_i(a) \geq g_i(b)+t_i(g_i(b))} p_i. \tag{6.14}$$

Note that preference thresholds, that lead to indifference zones, are used in a variant of the ELECTRE I method called ELECTRE IS (see Roy and Skalka (1984) or Roy and Bouyssou (1993)).

Thresholding is a key tool in the original outranking methods; it allows one to bypass the necessity of transforming the original evaluations to obtain linear scales. There is another occasion for invoking thresholds, which is in the analysis of *discordance*.

Discordance and vetoes

Remember that the principle of the outranking methods consists in examining the validity of the proposition "*a* outranks *b*"; the concordance index "measures" the arguments in favour of saying so, but there may be arguments strongly against that assertion (*discordant* criteria). These discordant voices can be viewed as vetoes; there is a veto against declaring that a outranks b if b is so much better than a on some criterion that it becomes disputable or even meaningless to pretend that a might be better overall than b. Let us emphasise that the effect of a veto is quite radical, just like in the voting context. If a veto threshold is passed on a criterion when comparing two alternatives, then the alternative against which there is a veto, say a, may not outrank the other one, say b; this may result in *incomparabilities* in the outranking relation if in addition b does not outrank a, either because the coalition of criteria stating that b is at least as good as a is not strong enough or because there is also a veto of a against b on another criterion.

To be more precise, a veto threshold on criterion i is in general a function v_i encoding a difference in evaluations so big that it would be out of the question to say that a outranks b if

$$g_i(a) > g_i(b) + v_i(g_i(b)) \tag{6.15}$$

when criterion i is to be minimised, or

$$g_i(a) < g_i(b) - v_i(g_i(b)) \tag{6.16}$$

when criterion i is to be maximised. Of course it may be the case that the function v_i be a constant.

In our case, in view of Thierry's particular interest in sporty cars, the criterion most likely to yield a veto is acceleration. Although there was no precise indication on setting vetoes in Thierry's preliminary analysis (section 6.1.2), one might speculate that on the acceleration criterion, pairs such as (28, 29.6), (28.3, 30), (29, 30.4), (29, 30.7) (all evaluations expressed in seconds) and all intervals wider than those listed, lead to a veto (against claiming that the alternative with the higher evaluation could be preferred to the other one, since here, the criterion is to be minimised). If this would seem reasonable then we would not be far from accepting a constant veto threshold of about 1.5 or 1.6 second. If we decide that there is a veto with a constant threshold on the *acceleration* criterion for differences exceeding 1.5 second, it means that a car that accelerates from 0 to 100 km/h in 29.6 seconds (as is the case of Peugeot 309 GTI) could not conceivably outrank a car which does it in 28 (as Honda Civic does) whatever the evaluations on the other criteria might be. Of course, setting the veto threshold to 1.5 implies that a car needing 30.4 seconds (like Mazda 323) may not outrank a car that accelerates in 28.9 (like Opel Astra or Renault 21) but might very well outrank a car that accelerates in 29 (like Nissan Sunny) if the performances on the other criteria are superior. Using 1.5 as a veto threshold thus implies that differences of at least 1.5 from 28 to 29.6 or from 28.9 to 30.4 have the same consequences in terms of preference. Setting the value of the veto threshold obviously involves some degree of arbitrariness; why not set the threshold at 1.4 second, which would imply that Mazda 323 may not outrank Nissan Sunny? In such cases, it must be verified whether small variations around the chosen value of a parameter (such as a veto threshold) do not influence the conclusions in a dramatic manner; if small variations do have a strong influence, detailed investigation is needed in order to decide which setting of the parameter's value is most appropriate. A related facet of using thresholds is that growing differences that are initially not significant, brutally crystallise into significant ones as soon as a crisp threshold is passed; obviously methods using thresholds may show discontinuities in their consequences and that is why sensitivity analysis is even more crucial here than with more classical methods. However, the underlying logic is quite similar to that on which statistical tests are based; here as well, conventional levels of significance (like the famous 5% rejection intervals) are widely used to decide whether a hypothesis must be rejected or not. We will allude in the next section to more "gradual" methods that can be designed on the basis of concordance-discordance principles similar to those outlined above.

In order not to be too long we do not develop the consequences of introducing veto thresholds in our example. It suffices to say that the outranking relation, its kernel and the derived rankings are not dramatically modified in the present case.

6.4.4 Main features and problems of elementary outranking approaches

The ideas behind the methods analysed above may be summarised as follows. For each pair of alternatives (a, b) it is determined whether a outranks b by comparing their evaluations $g_i(a)$ and $g_i(b)$ on each point of view i. The pairs of evaluations are compared to intervals that can be viewed as typical of classes of ordered pairs of evaluations on each criterion (for instance the classes "indifference", "preference" and "veto"). On the basis of the list of classes to which it belongs for each criterion (its "profile"), the pair (a, b) is declared to be or not to be in the outranking relation.

Note that

- a credibility index of outranking (for instance "weak" and "strong" outranking) may be defined; to each value of the index corresponds a set of profiles;

if the profile of the pair (a, b) is one of those associated with a particular value of credibility of outranking, then the outranking of b by a is assigned this value of credibility index; there are of course rationality requirements for the sets of profiles associated with the various values of the credibility index; this credibility index is to be interpreted in logical terms; it models the degree to which it is *true* that there are enough arguments in favour of saying that a is better than b while there is no strong reason of refuting this statement (see the definition of outranking in Section 6.4.2);

- thresholds may be used to determine the classes in differences for preference on each criterion, provided differences $g_i(a) - g_i(b)$ equal to such thresholds have the same meaning independently of their location on the scale of criterion i (linearity property);
- the rules for determining whether a outranks b (eventually to some degree of a credibility index) generally involve weights that describe the relative importance of the criteria; these weights are typically used additively to measure the importance of coalitions of criteria independently of the evaluations of the alternatives.

The result of the construction, i.e. the outranking relation (possibly qualified with a degree of a credibility index), is then exploited in view of a specific type of decision problems (choice, ranking, . . .). It is supposed to include all the relevant and sure information about preference that could be extracted from the data and the questions answered by the decision-maker.

Due to their lack of transitivity and acyclicity, procedures are needed to derive a ranking or a choice set from the outranking relation. In the process of deriving a complete ranking from the outranking relation, the property of independence of irrelevant alternatives (see Chapter 2 where this property is evoked) is lost; this property was satisfied in the construction of the outranking relation since outranking is decided by looking in turn at the profiles of each pair of alternatives, independently of the rest. Since this is an hypothesis of Arrow's theorem and it is violated, the conclusion of the theorem is not necessarily valid and one may hope that there is no criterion playing the role of dictator.

The various procedures that have been proposed for exploiting the outranking relation (for instance transforming it into a complete ranking) are not above criticism; it is especially difficult to justify them rigorously since they operate on an object that has been constructed, the outranking relation. Since the decision-maker has no direct intuition of this object, one can hardly expect to get reliable answers when questioning him about the properties of this relation. On the other hand, a direct characterisation of the ranking produced by the exploitation of an outranking relation seems out of reach.

Non-compensation

The weights count entirely or not at all in the comparison of two alternatives; the smaller or larger difference in evaluations between alternatives does not matter once a certain threshold is passed. This fact, which was discussed in the second paragraph of this section 6.4.3, is sometimes called the *non-compensation* property of outranking methods. A large difference in favour, say, of a over b on some criterion is of no use to compensate for small differences in favour of b on many criteria since all that counts for deciding that a outranks b is the list of criteria in favour of a. Vetoes only have a “negative” action, impeding that outranking be declared. The reader interested in the non-compensation property is referred to Fishburn (1976), Bouyssou and Vansnick (1986), Bouyssou (1986).

Incomparability and indifference

For some pairs (a, b) it may be the case that neither a outranks b nor the opposite; this can occur not only because of the activation of a veto but alternatively because the credibility of both the outranking of a by b and of b by a are not sufficiently high. In such a case a and b are said to be *incomparable*. This may be interpreted in two different ways. One may advance that some alternatives are too contrasted to be compared. It has been argued, for instance, that comparing a Rolls Royce with a small and cheap car proves impossible because the Rolls Royce is incomparably better on many criteria but is also incomparably more expensive. Another example concerns the comparison of projects that involve the risk of loss of human life; should one prefer a more expensive project with a lower risk or a less expensive one with higher risk (see Chapter 5, Section 5.3.3, for evaluations of the cost of human losses in various countries)? Other people support the idea that incomparability results from insufficient information; the available information sometimes does not allow to make up one’s mind on whether a is preferred to b or the converse.

In any case, incomparability should not be assimilated to indifference. Indifference occurs when alternatives are considered as almost equivalent; incomparability is more concerned with very contrasted alternatives. The treatment of the two categories is quite different in the exploitation phase; indifferent alternatives should appear in the same class of a ranking or in neighbouring one, while incomparable alternatives may be ranked in classes quite far apart.

6.4.5 Advanced outranking methods: from thresholding towards valued relations

Looking at the variants of the ELECTRE method suggests that there is a general pattern on which they are all built:

- alternatives are considered in pairs and eventually, outranking is determined on the basis of the profiles of performance of the pair only;
- the differences between the evaluations of a pair of alternatives for each criterion are categorised in discrete classes delimited by thresholds (preference, veto, ...);
- rules are invoked to decide which combinations of these classes lead to outranking; more generally, there are several grades of outranking (weak, strong in ELECTRE II, ...) and rules associate specific combinations of classes to each grade;

- specialised procedures are used to exploit the various grades of outranking in view of supporting the decision process.

Defining the classes through thresholding raises the problem of discontinuity alluded to in the previous section. It is thus appealing to work with continuous classes of differences of preference for each criterion, i.e. directly with valued relations. A value $c_j(a, b)$ on arc (a, b) models the degree to which alternative a is preferred to alternative b on criterion j. These degrees are often interpreted in logical fashion as a degree of credibility of the preference. Then each combination of values of the credibility index on the various criteria may be assigned an *overall* value of the credibility index for outranking; the outranking relation is also valued in such a context.

Dealing with valued relations and especially combining "values" raises a question: which operations may be meaningfully (or just reasonably) performed on them. Our analysis of the weighted sum in section 6.2 has taught us that operations that may appear as natural, rely on strong assumptions that suppose very detailed information on the preferences.

Consider the following formula which is used in ELECTRE III, a method leading to a valued outranking relation (see Roy and Bouyssou (1993) or Vincke (1992)), to compute the overall degree of credibility $S(a, b)$ of the outranking of b by a.

$$S(a,b) = \begin{cases} c(a,b) & \text{if } D_j(a,b) \leq c(a,b)\ \forall j \\ c(a,b) \times \prod_{j: D_j(a,b) > c(a,b)} \frac{1 - D_j(a,b)}{1 - c(a,b)} & \text{otherwise} \end{cases}$$

In the above formula, $D_j(a, b)$ is a degree of credibility of discordance. We do not enter into the detail of how $c(a, b)$ or $D_j(a, b)$ can be computed; just remember that they are valued between 0 and 1.

The justification of such a formula is mainly heuristic in the sense that the response of the formula to the variation of some inputs is not counter-intuitive: when discordance raises outranking decreases; the converse with concordance; when discordance is maximal there may not be any degree of outranking at all. This does not mean that the formula is fully justified. Other formulae might have been chosen with similarly good heuristic behaviour. The weighted sum also has good heuristic properties at first glance, but deeper investigation shows that the values it yields cannot be trusted as a valid representation of the preferences unless additional information is requested from the decision-maker and used to re-code the original evaluations g_j. The formula above involves operations such as multiplication and division that suppose that concordance and discordance indices are plainly cardinal numbers and not simply labels of ordered categories. This is indeed a strong assumption that does not seem to us to be supported by the rest of the approach, in particular by the manner in which the indices are elaborated; in the elementary outranking methods (ELECTRE I and II) much care was taken, for instance, to avoid performing arithmetical operations on the evaluations $g_i(a)$; only cuts of the concordance index were considered (which is typically an operation valid for ordinal data); vetoes were used in a very radical fashion. No special attention, comparable to what was needed to build value functions from the evaluations, was paid to building concordance and discordance indices; in particular, nothing guarantees that these indices can be combined by means of arithmetic operations and produce an overall index S representative of a degree of credibility of an outranking. For instance, consider the following two cases which lead to an outranking degree of .4:

- the concordance index $c(a, b)$ is equal to .40 and there is no discordance (i.e. $D_j(a, b) = 0$ for all j);
- the concordant coalition weighs .80 but there is a strong discordance on criterion 1; $D_1(a, b) = .90$ while $D_j(a, b) = 0$ for all $j \neq 1$.

For both, the formula yields a degree of outranking of .40. Obviously another formula with similar heuristic behaviour might have resulted in quite different outputs. Consider for instance the following:

$$S(a, b) = \min\{c(a, b), \min\{1 - D_j(a, b), j = 1, \dots, n\}\}$$

On the first case, it yields an outranking degree of .40 as well but on the second case, the degree falls to .10. It is likely that in some circumstances a decision-maker might find the latter model more appropriate. Note also that the latter formula does not involve arithmetic operations on $c(a, b)$ and the $1 - D_j(a, b)$'s but only ordinal operations, namely taking the minimum. This means that transforming $c(a, b)$ and the $1 - D_j(a, b)$'s by an increasing transformation of the $[0, 1]$ interval would just amount to transforming the original value of $S(a, b)$ by the same transformation. This is not the case with the former formula. Hence, if the information content of the $c(a, b)$ and the $1 - D_j(a, b)$'s just consists in the ordering of their values in the $[0, 1]$ interval, then the former formula is not suitable. For a survey of possible ways of aggregating preferences into a valued relation, the reader is referred to chapters 2 and 3 of the book edited by Słowiński (1998).

The fact that the value obtained for the outranking degree may involve some degree of arbitrariness did not escape Roy and Bouyssou (1993) who explain (p.417) that the value of the degree of outranking obtained by a formula like the above should be handled with care; they advocate that thresholds be used when comparing two such values: the outranking of b by a can be considered to be more credible than the outranking of d by c only if $S(a, b)$ is significantly larger than $S(c, d)$. We agree with this statement but unfortunately it seems quite difficult to assign a value to a threshold above which the difference $S(a, b) - S(c, d)$ could be claimed as "significant".

There are thus two directions that can be followed for taking the objections to the formula of ELECTRE III into account. In the first option, one considers that the meaning of the concordance and discordance degrees is ordinal and one tries to determine a family of aggregation formulae that fulfil basic requirements including compatibility with the ordinal character of concordance and discordance. The other option consists in revising the way concordance and discordance indices are constructed in order to have a quantitative meaning that allows to use arithmetic operations for aggregating them. That is, at least tentatively, the option followed in the PROMETHEE methods (see Brans and Vincke (1985) or Vincke (1992); these methods may be interpreted as aiming towards building a value function on the pairs of alternatives; this function would represent the overall difference in preference between any two alternatives. The way that this function is constructed in practice however, leaves the door open to remarks analogous to those addressed to the weighted sum in Section 6.2.

6.5 General conclusion

This long chapter has enabled us to travel through the continent of formal methods of decision analysis; by "formal" we mean those methods relying on an explicit mathematical model of the decision-maker's preferences. We neither looked into all methods nor did we explore those we looked into completely. There are other continents that have been almost completely ignored, in particular all the methods that do not rely on a formal modelling of the preferences (see for instance the book edited by Rosenhead (1989) in which various approaches are presented for structuring problems in view of facilitating decision making).

On the particular topic of multi-attribute decision analysis, we may summarise our main conclusions as follows:

- Numbers do not always mean what they seem to. It makes no sense to manipulate raw evaluations without taking the context into account. Numbers may have an ordinal meaning, in which case it cannot be recommended to perform arithmetic operations on them; they may be evaluations on an interval scale or a ratio scale and there are appropriate transformations that are allowed for each type of scale. We have also suggested that the significance of a number may be intermediate between ordinal and cardinal; in that case, the interval separating two evaluations might be given an interpretation: one might take into consideration the fact that intervals are e.g. large, medium or small. Evaluations may also be imprecise and knowing that should influence the way they will be handled. *Preference modelling* is specifically the activity that deals with the meaning of the data in a decision context.

- Preference modelling does not only take objective information linked with the evaluations or with the data, such as the type of scale or the degree of precision or the degree of certainty into account. It also incorporates subjective information in relation to the preferences of the decision maker. Even if numeric evaluations actually mean what they seem to, their significance is not immediately in terms of preferences: the interval separating two evaluations must be reinterpreted in terms of difference in preferences.

- The (vague) notion of *importance* of the criteria and its implementation are strongly model-dependent. Weights and trade-offs should not be elicited in the same manner depending on the type of model since e.g. they may or may not depend on the scaling of the criteria.

- There are various types of models that can be used in a decision process. There is no best model; all have their strong points and their weak points.

The choice of a particular approach (including a type of model) should be the result of an evaluation, in a given decision situation, of the chances of being able to elicit the parameters of the corresponding model in a reliable manner; these "chances" obviously depend on several factors including the type and precision of the available data, the way of thinking of the decision-maker, his knowledge of the problem. Another factor that should be considered for choosing a model, is the type of information that is wanted as output: the decision maker needs different information when he has to rank alternatives to when he has to choose among alternatives or when he has to assign them to predefined (ordered) categories (we put the latter problem aside in our discussion of the car choosing case). So, in our view, the ideal decision analyst, should master several methodologies for building a model. Notice that additional dimensions make the choice and the construction of a model in group decision making even more difficult; the dynamics of such decision processes is by far more complex, involving conflicts and negotiation aspects; constructing complete formal models in such contexts is not always possible, but it remains that using problem structuring tools (such as cognitive maps) may prove profitable.

- A direct consequence of the possibility of using different models is that the output may be discordant or even contradictory. We have encountered such a situation several times in the above study; cars may be ranked in different positions according to the method that is used. This does not puzzle us too much. First of all, because the observed differences appear more as variants than as contradictions; the various outputs are remarkably consistent and the variants can be explained to some extent. Second, the approaches use different concepts and the questions the decision maker has to answer are accordingly expressed in different languages; this of course induces variability. This is no wonder since the information that decision analysis aims at capturing cannot usually be precisely measured. It is sufficient to recall that experiments have shown that there is much variability in the answers of subjects submitted to the same questions at time intervals. Does this mean that all methods are acceptable? Not at all. There are several criteria of validity. One is that the method has to be accepted in a particular decision situation; this means that the questions asked to the decision-maker must make sense to him and he should not be asked for information he is unable to provide in a reliable manner. There are also internal and external consistency criteria that a method should fulfil. Internal consistency implies making explicit the hypotheses under which data form an acceptable input for a method; then the method should perform operations on the input that are compatible with the supposed properties of the input; this in turn induces an output which enjoys particular properties. External consistency consists in checking whether the available information matches the requirements of acceptable inputs and whether the output may help in the decision process. The main goal of the above study was to illustrate the issue of internal and external validity on a few methods in a specific simple problem.

Besides the above points that are specific to multiple criteria preference models, more general lessons can also be drawn.

- If we consider our trip from the weighted sum to the additive multi-attribute value model in retrospect, we see that much self-confidence and therefrom much convincing power can be gained by eliciting conditions under which an approach such as the weighted sum would be legitimate. The analysis is worth the effort because precise concepts (like trade-offs and values) are sculptured through analysis that also results in methods for eliciting the parameters of the model. Another advantage of theory is to provide us with limits, i.e. conditions under which a model is valid and a method is applicable. From this viewpoint and although the outranking methods have not been fully characterised, it is worth noticing that their study has recently made theoretical progress (see e.g. Arrow and Raynaud (1986), Bouyssou and Perny (1992), Vincke (1992), Fodor and Roubens (1994), Tsoukiàs and Vincke (1995) , Bouyssou (1996), Marchant (1996), Bouyssou and Pirlot (1997)), Pirlot (1997)) .
- An advantage of formal models that could not be overemphasised is that they favour communication. In the course of the decision process, the construction of the model requires that pieces of information, knowledge and priorities that are usually implicit or hidden, be brought into light and taken into account; also, the choice of the model reflects the type of available information (more or less certain, precise, quantitative). The result is often a synthesis of what is known and what has been learnt about the decision problem in the process of elaborating the model. The fact that a model is formal also allows for some sort of calculations; in particular, testing to what extent the conclusions are stable when the evaluation of imprecise data are varied is possible within formal models. Once a decision has been made, the model does not lose its utility. It can provide grounds for arguing in favour or against a decision. It can be adapted to make ulterior decisions in similar contexts.
- The "decisiveness" of the output depends on the "richness" of the information available. If the knowledge is uncertain, imprecise or simply non-quantitative in nature, it may be difficult to build a very strong model; by "strong", we mean a model that clearly suggests a decision as, for instance, those that produce a ranking of the alternatives. Other models (and especially those based on pairwise comparisons of alternatives and verifying the independence of irrelevant alternatives property) are not able—structurally—to produce a ranking; they may nevertheless be the best possible synthesis of the relevant information in particular decision situations. In any case, even if the model leads to a ranking, the decision is to be taken by the decision-maker and it is not in general an automatic consequence of the model (due for instance to imprecisions in the data that calls for a relativisation of the model's prescription). As will be illustrated in greater detail in Chapter 9, the construction of a model is not all of the decision process.

Bibliography

[1] Abbas, M., Pirlot, M. and Vincke, Ph. (1996). Preference structures and co-comparability graphs, *Journal of Multicriteria Decision Analysis* **5**: 81–98.

[2] Abdellaoui, M. and Munier, B. (1994). The 'closing in' method: An experimental tool to investigate individual choice patterns under risk, *in* B. Munier and M.J. Machina (eds), *Models and experiments in risk and rationality*, Kluwer, Dordrecht, pp. 141–155.

[3] Adler, H.A. (1987). *Economic appraisal of transport projects: A manual with case studies*, Johns Hopkins University Press for the World Bank, Baltimore.

[4] Airaisian, P.W. (1991). *Classroom assessment*, McGraw-Hill, New York.

[5] Allais, M. and Hagen, O. (eds) (1979). *Expected utility hypotheses and the Allais paradox*, D. Reidel, Dordrecht.

[6] Allais, M. (1953). Le comportement de l'homme rationnel devant le risque : Critique des postulats et axiomes de l'école américaine, *Econometrica* **21**: 503–46.

[7] Armstrong, W.E. (1939). The determinateness of the utility function, *The Economic Journal* **49**: 453–467.

[8] Arrow, K.J. and Raynaud, H. (1986). *Social choice and multicriterion decision-making*, MIT Press, Cambridge.

[9] Arrow, K.J. (1963). *Social choice and individual values*, 2nd edn, Wiley, New York.

[10] Atkinson, A.B. (1970). On the measurement of inequality, *Journal of Economic Theory* **2**: 244–263.

[11] Baldwin, J.F. (1979). A new approach to approximate reasoning using a fuzzy logic, *Fuzzy Sets and Systems* **2**: 309–325.

[12] Balinski, M.L. and Young, H.P. (1982). *Fair representation*, Yale University Press, New Haven.

[13] Bana e Costa, C.A., Ensslin, L., Corrêa, E.C. and Vansnick, J.-C. (1999). Decision support systems in action: Integrated application in a multicriteria decision aid process, *European Journal of Operational Research* **113**: 315–335.

[14] Barbera, S., Hammond, P. and Seidl, C. (eds) (1998). *Handbook of utility theory*, Vol. 1: *Principles*, Kluwer, Dordrecht.

[15] Bartels, R. H.., Beatty, J. C.. and Barsky, B.H.. (1987). *An introduction to Spline for use in computer graphics and geometric Modeling*, Morgan Kaufmann, Los Altos.

[16] Barzilai, J., Cook, W.D. and Golany, B. (1987). Consistent weights for judgments matrices of the relative importance of alternatives, *Operations Research Letters* **6**: 131–134.

[17] Bazerman, M.H. (1990). *Judgment in managerial decision making*, Wiley, New York.

[18] Bell, D., Raiffa, H. and Tversky, A. (eds) (1988). *Decision making: Descriptive, normative and prescriptive interactions*, Cambridge University Press, Cambridge.

[19] Belton, V., Ackermann, F. and Shepherd, I. (1997). Integrated support from problem structuring through alternative evaluation using COPE and V•I•S•A, *Journal of Multi-Criteria Decision Analysis* **6**: 115–130.

[20] Belton, V. and Gear, A.E. (1983). On a shortcoming of Saaty's analytic hierarchies, *Omega* **11**: 228–230.

[21] Belton, V. (1986). A comparison of the analytic hierarchy process and a simple multi-attribute value function, *European Journal of Operational Research* **26**: 7–21.

[22] Béreau, M. and Dubuisson, B. (1991). A fuzzy extended k-nearest neighbor rule, *Fuzzy Sets and Systems* **44**: 17–32.

[23] Bernoulli, D. (1954). Specimen theoriæ novæ de mensura sortis, *Commentarii Academiæ Scientiarum Imperialis Petropolitanæ* (**5**, 175–192, 1738), *Econometrica* **22**: 23–36. Translated by L. Sommer.

[24] Bezdek, J., Chuah, S.K. and Leep, D. (1986). Generalised k-nearest neighbor rules, *Fuzzy Sets and Systems* **18**: 237–256.

[25] Blin, M.-J. and Tsoukiàs, A. (1998). Multicriteria methodology contribution to the software quality evaluation, *Technical report*, Cahier du LAMSADE N^o 155, Université Paris-Dauphine, Paris.

[26] Boardman, A. (1996). *Cost benefit analysis: Concepts and practices*, Prentice-Hall, New-York.

[27] Boiteux, M. (1994). *Transports : Pour un meilleur choix des investissements*, La Documentation Française, Paris.

[28] Bonboir, A. (1972). *La docimologie*, PUF, Paris.

[29] Borda, J.-Ch. (1781). *Mémoire sur les élections au scrutin*, Comptes Rendus de l'Académie des Sciences. Translated by Alfred de Grazia as "Mathematical derivation of an election system", *Isis*, Vol. **44**, pp. 42–51.

[30] Bouchon, B. (1995). *La logique floue et ses applications*, Addison Wesley, New York.

[31] Bouchon-Meunier, B. and Marsala, C. (1999). Learning fuzzy decision rules, *in* D. D. J. Bezdek and H. Prade (eds), *Fuzzy sets in approximate reasoning and information systems*, Vol. 3 of *Handbook of Fuzzy Sets*, Kluwer, Dordrecht, chapter 4, pp. 279–304.

[32] Bouyssou, D., Perny, P., Pirlot, M., Tsoukiàs, A. and Vincke, Ph. (1993). A manifesto for the new MCDM era, *Journal of Multi-Criteria Decision Analysis* **2**: 125–127.

[33] Bouyssou, D. and Perny, P. (1992). Ranking methods for valued preference relations: A characterization of a method based on entering and leaving flows, *European Journal of Operational Research* **61**: 186–194.

[34] Bouyssou, D. and Pirlot, M. (1997). Choosing and ranking on the basis of fuzzy preference relations with the 'Min in Favor', *in* G. Fandel and T. Gal (eds), *Multiple criteria decision making – Proceedings of the twelfth international conference, Hagen, Germany*, Springer Verlag, Berlin, pp. 115–127.

[35] Bouyssou, D. and Vansnick, J.-C. (1986). Noncompensatory and generalized noncompensatory preference structures, *Theory and Decision* **21**: 251–266.

[36] Bouyssou, D. (1984). Decision-aid and expected utility theory: A critical survey, *in* O. Hagen and F. Wenstøp (eds), *Progress in utility and risk theory*, Kluwer, Dordrecht, pp. 181–216.

[37] Bouyssou, D. (1986). Some remarks on the notion of compensation in MCDM, *European Journal of Operational Research* **26**: 150–160.

[38] Bouyssou, D. (1990). Building criteria: A prerequisite for MCDA, *in* C.A. Bana e Costa (ed.), *Readings in multiple criteria decision aid*, Springer Verlag, Berlin, pp. 58–80.

[39] Bouyssou, D. (1992). On some properties of outranking relations based on a concordance-discordance principle, *in* A. Goicoechea, L. Duckstein and S. Zionts (eds), *Multiple criteria decision making*, Springer-Verlag, Berlin, pp. 93–106.

[40] Bouyssou, D. (1996). Outranking relations: Do they have special properties?, *Journal of Multi-Criteria Decision Analysis* **5**: 99–111.

[41] Brams, S.J. and Fishburn, P.C. (1982). *Approval voting*, Birkhäuser, Basel.

[42] Brans, J.-P. and Vincke, Ph. (1985). A preference ranking organization method, *Management Science* **31**: 647–656.

[43] Brekke, K.A. (1997). The numéraire matters in cost-benefit analysis, *Journal of Public Economics* **64**: 117–123.

[44] Brent, R.J. (1984). Use of distributional weights in cost-benefit analysis: A survey of schools, *Public Finance Quarterly* **12**: 213–230.

[45] Brent, R.J. (1996). *Applied cost-benefit analysis*, Elgar, Adelshot Hants.

[46] Broome, J. (1985). The economic value of life, *Economica* **52**: 281–294.

[47] Carbone, E. and Hey, J.D. (1995). A comparison of the estimates of expected utility and non-expected utility preference functionals, *Geneva Papers on Risk and Insurance Theory* **20**: 111–133.

[48] Cardinet, J. (1986). *Évaluation scolaire et mesure*, De Boeck, Brussels.

[49] Chatel, E. (1994). Qu'est-ce qu'une note : recherche sur la pluralité des modes d'éducation et d'évaluation, *Les Dossiers d'Éducation et Formations* **47**: 183–203.

[50] Checkland, P. (1981). *Systems thinking, systems practice*, Wiley, New York.

[51] Condorcet, M.J.A.N.C., marquis de. (1785). *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*, Imprimerie Royale, Paris.

[52] Cover, T. M. and Hart, P. E. (1967). Nearest neighbor pattern classification, *IEEE, Transactions on Information Theory, IT-13* **1**: 21–27.

[53] Cross, L.H. (1995). Grading students, *Technical Report Series EDO-TM-95-5*, ERIC/AE Digest.

[54] Daellenbach, H.G. (1994). *Systems and decision making. A management science approach*, Wiley, New York.

[55] Dasgupta, P.S., Marglin, S. and Sen, A.K. (1972). *Guidelines for project evaluation*, UNIDO, New York.

[56] Dasgupta, P.S. and Pearce, D.W. (1972). *Cost-benefit analysis: Theory and practice*, Macmillan, Basingstoke.

[57] Davis, B.G. (1993). *Tools for teaching*, Jossey-Bass, San Francisco.

[58] de Jongh, A. (1992). *Théorie du mesurage, agrégation des critères et application au décathlon*, Master's thesis, SMG, Université Libre de Bruxelles, Brussels.

[59] Dekel, E. (1986). An axiomatic characterization of preference under uncertainty: Weakening the independence axiom, *Journal of Economic Theory* **40**: 304–318.

[60] Desrosières, A. (1995). Refléter ou instituer : L'invention des indicateurs statistiques, *Technical Report 129/J310*, INSEE, Paris.

[61] de Ketele, J.-M. (1982). *La docimologie*, Cabay, Louvain-La-Neuve.

[62] de Landsheere, G. (1980). *Évaluation continue et examens. Précis de docimologie*, Labor-Nathan, Paris.

[63] Dinwiddy, C. and Teal, F. (1996). *Principles of cost-benefit analysis for developing countries*, Cambridge University Press, Cambridge.

[64] Dorfman, R. (1996). Why benefit-cost analysis is widely disregarded and what to do about it?, *Interfaces* **26**: 1–6.

[65] Drèze, J. and Stern, N. (1987). The theory of cost-benefit analysis, *in* A.J. Auebach and M. Feldstein (eds), *Handbook of public economics*, Elsevier, Amsterdam, pp. 909–989.

[66] Dubois, D., Fargier, H. and Prade, H. (1997). Decision-making under ordinal preferences and uncertainty, *in* D. Geiger and P.P. Shenoy (eds), *Proceedings of the 13th conference on uncertainty in artificial intelligence*, Morgan Kaufmann, Los Altos, pp. 157–164.

[67] Dubois, D., Prade, H. and Sabbadin, R. (1998). Qualitative decision theory with Sugeno integrals, *Proceedings of the $14^t h$ conference on uncertainty in artificial intelligence*, Morgan Kaufmann, Los Altos, pp. 121–128.

[68] Dubois, D., Prade, H. and Ughetto, L. (1999). Fuzzy logic, control engineering and artificial intelligence, *in* H.B. Verbruggen, H.J. Zimmermann and R. Babuska (eds), *Fuzzy algorithms for control*, Kluwer, Dordrecht, pp. 17–58.

[69] Dubois, D. and Prade, H. (1987). The mean value of a fuzzy number, *Fuzzy Sets and Systems* **24**: 279–300.

[70] Dubois, D. and Prade, H. (1988). *Possibility theory*, Plenum Press, New-York.

[71] Dupuit, J. (1844). De la mesure de l'utilité des travaux publics, *Annales des Ponts et Chaussées* (8).

[72] Dyer, J.S. (1990). Remarks on the analytic hierarchy process, *Management Science* **36**: 249–258.

[73] Ebel, R.L. and Frisbie, D.A. (1991). *Essentials of educational measurement*, Prentice-Hall, New-York.

[74] Ellsberg, D. (1961). Risk, ambiguity and the Savage axioms, *Quarterly Journal of Economics* **75**: 643–669.

[75] Fargier, H. and Perny, P. (1999). Qualitative decision models under uncertainty without the commensurability hypothesis, *in* K.B.. Laskey and H. Prade (eds), *Proceedings of the $15^t h$ conference on uncertainty in artificial intelligence*, Morgan Kaufmann, Los Altos, pp. 188–195.

[76] Farrell, D.M. (1997). *Comparing electoral systems*, Contemporary Political Studies, Prentice-Hall, New-York.

[77] Fiammengo, A., Buosi, D., Iob, I., Maffioli, P., Panarotto, G. and Turino, M. (1997). Bid management of software acquisition for cartography applications. Presented at AIRO '97 Conference, Aosta.

[78] Fishburn, P.C. and Sarin, R.K. (1991). Dispersive equity and social risk, *Management Science* **37**: 751–769.

[79] Fishburn, P.C. and Sarin, R.K. (1994). Fairness and social risk I: Unaggregated analyses, *Management Science* **40**: 1174–1188.

[80] Fishburn, P.C. and Straffin, P.D. (1989). Equity considerations in public risks evaluation, *Operations Research* **37**: 229–239.

[81] Fishburn, P.C. (1970). *Utility theory for decision-making*, Wiley, New York.

[82] Fishburn, P.C. (1976). Noncompensatory preferences, *Synthese* **33**: 393–403.

[83] Fishburn, P.C. (1977). Condorcet social choice functions, *SIAM Journal on Applied Mathematics* **33**: 469–489.

[84] Fishburn, P.C. (1978). A survey of multiattribute/multicriteria evaluation theories, *in* S. Zionts (ed.), *Multicriteria problem solving*, Springer Verlag, Berlin, pp. 181–224.

[85] Fishburn, P.C. (1982). *The foundations of expected utility*, D. Reidel, Dordrecht.

[86] Fishburn, P.C. (1984). Equity axioms for public risks, *Operations Research* **32**: 901–908.

[87] Fishburn, P.C. (1988a). *Nonlinear preference and utility theory*, Johns Hopkins University Press, Baltimore.

[88] Fishburn, P.C. (1988b). Normative theories of decision making under risk and under uncertainty, *in* M. Kacprzyk and M. Roubens (eds), *Nonconventional preference relations in decision making*, Springer Verlag, Berlin, pp. 469–489.

[89] Fishburn, P.C. (1991). Nontransitive preferences in decision theory, *Journal of Risk and Uncertainty* **4**: 113–134.

[90] Fix, E. and Hodges, J.L. (1951). Discriminatory analysis, non-parametric discrimination: consistency properties, *Technical report*, USAF Scholl of aviation and medicine, Randolph Field. 4.

[91] Fodor, J.C. and Roubens, M. (1994). *Fuzzy preference modelling and multicriteria decision support*, Kluwer, Dordrecht.

[92] Folland, S., Goodman, A.C. and Stano, M. (1997). *The economics of health and health care*, Prentice-Hall, New-York.

[93] French, S. (1981). Measurement theory and examinations, *British Journal of Mathematical and Statistical Psychology* **34**: 38–49.

[94] French, S. (1993). *Decision theory – An introduction to the mathematics of rationality*, Ellis Horwood, London.

[95] Gacogne, L. (1997). *Eléments de logique floue*, Hermès, Paris.

[96] Gafni, A. and Birch, S. (1997). Equity considerations in utility-based measures of health outcomes in economic appraisals: An adjustment algorithm, *Journal of Health Economics* **10**: 329–342.

[97] Gehrlein, W.V. (1983). Condorcet's paradox, *Theory and Decision* **15**: 161–197.

[98] Gibbard, A. (1973). Manipulation of voting schemes: A general result, *Econometrica* **41**: 587–601.

[99] Gilboa, I. and Schmeidler, D. (1989). Maxmin expected utility with a non-unique prior, *Journal of Mathematical Economics* **18**: 141–153.

[100] Gilboa, I. and Schmeidler, D. (1993). Updating ambigous beliefs, *Journal of Economic Theory* **59**: 33–49.

[101] Grabisch, M., Guely, F. and Perny, P. (1997). *Évaluation subjective*, Les cahiers du Club CRIN - Association ECRIN, Paris.

[102] Grabisch, M. (1996). The application of fuzzy integrals to multicriteria decision making, *European Journal of Operational Research* **89**: 445–456.

[103] Hammond, P.J. (1988). Consequentialist foundations for expected utility, *Theory and Decision* **25**: 25–78.

[104] Hanley, N. and Spash, C.L. (1993). *Cost-benefit analysis and the environment*, Elgar, Adelshot Hants.

[105] Harker, P.T. and Vargas, L.G. (1987). The theory of ratio scale estimation: Saaty's analytic hierarchy process, *Management Science* **33**: 1383–1403.

[106] Harless, D. and Camerer, C.F. (1994). The utility of generalized expected utility theories, *Econometrica* **62**: 1251–1289.

[107] Harvey, C.M. (1992). A slow-discounting model for energy conservation, *Interfaces* **22**: 47–60.

[108] Harvey, C.M. (1994). The reasonableness of non-constant discounting, *Journal of Public Economics* **53**: 31–51.

[109] Harvey, C.M. (1995). Proportional discounting of future costs and benefits, *Mathematics of Operations Research* **20**: 381–399.

[110] Henriet, L. and Perny, P. (1996). Méthodes multicrit res non-compensatoires pour la classification floue d'objets, *Proceedings of LFA'96*, pp. 9–15.

[111] Henriet, L. (1995). *Probl mes d'affectation et méthodes de classification*, Mémoire du dea 103, Université Paris Dauphine.

[112] Hershey, J.C., Kunreuther, H.C. and Schoemaker, P.J.H. (1982). Sources of bias in assessment procedures for utility functions, *Management Science* **28**: 936–953.

[113] Heurgon, E. (1982). Relationships between decision making process and study process in OR interventions, *European Journal of Operational Research* **10**: 230–236.

[114] Hey, J.D. and Orme, C. (1994). Investigating generalizations of expected utility theory using experimental data, *Econometrica* **62**: 1251–1289.

[115] Hogarth, R. (1987). *Judgement and choice: The psychology of decision*, Wiley, New York.

[116] Holland, A. (1995). The assumptions of cost-benefit analysis: A philosopher's view, *in* K.G. Willis and J.T. Corkindale (eds), *Environmental valuation: New perspectives*, CAB International, Oxford, pp. 21–38.

[117] Horn, R.V. (1993). *Statistical indicators*, Cambridge University Press, Cambridge.

[118] Humphreys, P.C., Svenson, O. and Vári, A. (1993). *Analysis and aiding decision processes*, North-Holland, Amsterdam.

[119] IEEE 92 (1992). Standard for a software quality metrics methodology, *Technical report*, The Institute of Electrical and Electronics Engineers.

[120] International Atomic Energy Agency (1993). *Cost-benefit aspects of food irradiation processing*, Bernan Associates, Washington D.C.

[121] ISO/IEC 9126 (1991). Information technology – Software product evaluation, quality characteristics and guidelines for their use, *Technical report*, ISO, Genève.

[122] Jacquet-Lagrèze, E., Moscarola, J., Roy, B. and Hirsch, G. (1978). Description d'un processus de décision, *Technical report*, Cahier du LAMSADE N^o 13, Université Paris-Dauphine, Paris.

[123] Jacquet-Lagrèze, E. and Siskos, J. (1982). Assessing a set of additive utility functions for multicriteria decision making: The UTA method, *European Journal of Operational Research* **10**: 151–164.

[124] Jacquet-Lagrèze, E. (1990). Interactive assessment of preferences using holistic judgments. The PREFCALC system, *in* C.A. Bana e Costa (ed.), *Readings in multiple criteria decision aid*, Springer Verlag, Berlin, pp. 335–350.

[125] Jaffray, J.-Y. (1988). Choice under risk and the security factor: An axiomatic model, *Theory and Decision* **24**: 169–200.

[126] Jaffray, J.-Y. (1989a). Some experimental findings on decision making under risk and their implications, *European Journal of Operational Research* **38**: 301–306.

[127] Jaffray, J.-Y. (1989b). Utility theory for belief functions, *Operations Research Letters* **8**: 107–112.

[128] Johannesson, M. (1995a). A note on the depreciation of the societal perspective in economic evaluation in health care, *Health Policy* **33**: 59–66.

[129] Johannesson, M. (1995b). The relationship between cost-effectiveness analysis and cost-benefit analysis, *Social Science and Medicine* **41**: 483–489.

[130] Johannesson, M. (1996). *Theory and methods of economic evaluation of health care*, Kluwer, Dordrecht.

[131] Johansson, P.O. (1993). *Cost-benefit analysis of environmental change*, Cambridge University Press, Cambridge.

[132] Johnson, E.J. and Schkade, D.A. (1989). Bias in utility assesments: Further evidence and explanations, *Management Science* **35**: 406–424.

[133] Kahneman, D., Slovic, P. and Tversky, A. (1981). *Judgement under uncertainty – Heuristics and biases*, Cambridge University Press, Cambridge.

[134] Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk, *Econometrica* **47**: 263–291.

[135] Keeler, E.B. and Cretin, S. (1983). Discounting of life-saving and other nonmonetary effects, *Management Science* **29**: 300–306.

[136] Keeney, R.L., Hammond, J.S. and Raiffa, H. (1999). *Smart choices: A guide to making better decisions*, Harvard University Press, Boston.

[137] Keeney, R.L. and Raiffa, H. (1976). *Decisions with multiple objectives: Preferences and value tradeoffs*, Wiley, New York.

[138] Keller, J., Gray, M. and Givens, J. (1985). A fuzzy $k-$nearest neighbor algorithm, *IEEE Transactions on Systems Man and Cybernetics.* **15**: 580–585.

[139] Kelly, J.S. (1991). Social choice bibliography, *Social Choice and Welfare* **8**: 97–169.

[140] Kerlinger, F.N. (1986). *Foundations of behavioral research*, 3rd edn, Holt, Rinehart and Winston, New York.

[141] Kirkpatrick, C. and Weiss, J. (1996). *Cost-benefit analysis and project appraisal in developing countries*, Elgar, Adelshot Hants.

[142] Kohli, K.N. (1993). *Economic analysis of investment projects: A practical approach*, Oxford University Press for the Asian Development Bank, Oxford.

[143] Krantz, D.H., Luce, R.D., Suppes, P. and Tversky, A. (1971). *Foundations of measurement*, Vol. 1: *Additive and polynomial representations*, Academic Press, New York.

[144] Krutilla, J.V. and Eckstein, O. (1958). *Multiple purpose river development*, Johns Hopkins University Press, Baltimore.

[145] Laska, J.A. and Juarez, T. (1992). *Grading and marking in American schools: Two centuries of debate*, Charles C. Thomas, Springfield.

[146] Laslett, R. (1995). The assumptions of cost-benefit analysis, *in* K.G. Willis and J.T. Corkindale (eds), *Environmental valuation: New perspectives*, CAB International, Oxford, pp. 5–20.

[147] Lesourne, J. (1975). *Cost-benefit analysis and economic theory*, North-Holland, Amsterdam.

[148] Lindheim, E., Morris, L.L. and Fitz-Gibbon, C.T. (1987). *How to measure performance and use tests*, Sage Publications, Thousand Oaks.

[149] Little, I.M.D. and Mirlees, J.A. (1968). *Manual of industrial project analysis in developing countries*, O.E.C.D, Paris.

[150] Little, I.M.D. and Mirlees, J.A. (1974). *Project appraisal and planning for developing countries*, Basic books, New York.

[151] Loomes, G. and Sugden, R. (1982). Regret theory: An alternative theory of rational choice under uncertainty, *Economic Journal* **92**: 805–824.

[152] Loomes, G. (1988). Different experimental procedures for obtaining valuations of risky actions: Implications for utility theory, *Theory and Decision* **25**: 1–23.

[153] Loomis, J., Peterson, G., Champ, P., Brown, T. and Lucero, B. (1998). Paired comparisons estimates of willingness to accept and contingent valuation estimates of willingness to pay, *Journal of Economic Behavior and Organisation* **35**: 501–515.

[154] Luce, R.D., Krantz, D.H., Suppes, P. and Tversky, A. (1990). *Foundations of measurement*, Vol. 3: *Representation, axiomatisation and invariance*, Academic Press, New York.

[155] Luce, R.D. and Raiffa, H. (1957). *Games and Decisions*, Wiley, New York.

[156] Luce, R.D. (1956). Semiorders and a theory of utility discrimination, *Econometrica* **24**: 178–191.

[157] Lysne, A. (1984). Grading of student's attainement: Purposes and functions, *Scandinavian Journal of Educational Research* **28**: 149–165.

[158] Machina, M.J. (1982). Expected utility without the independence axiom, *Econometrica* **50**: 277–323.

[159] Machina, M.J. (1989). Dynamic consistency and non-expected utility models of choice under uncertainty, *Journal of Economic Literature* **27**: 1622–1688.

[160] Mamdani, E. H.. (1981). *Gaines fuzzy reasonning and its applications*, Academic Press, New York.

[161] Marchant, Th. (1996). Valued relations aggregation with the Borda method, *Journal of Multi-Criteria Decision Analysis* **5**: 127–132.

[162] Masser, I. (1983). The representation of urban planning-processes: An exploratory review, *Environment and Planning B* **10**: 47–62.

[163] May, K.O. (1952). A set of independent necessary and sufficient conditions for simple majority decisions, *Econometrica* **20**: 680–684.

[164] McClennen, E.F. (1990). *Rationality and dynamic choice: Foundational explorations*, Cambridge University Press, Cambridge.

[165] McCord, M. and de Neufville, R. (1983). Fundamental deficiency of expected utility analysis, *in* S. French, R. Hartley, L.C. Thomas and D.J. White (eds), *Multiobjective decision making*, Academic Press, London, pp. 279–305.

[166] McCord, M. and de Neufville, R. (1982). Empirical demonstration that expected utility decision analysis is not operational, *in* B. Stigum and F. Wenstøp (eds), *Foundations of utility and risk theory*, D. Reidel, Dordrecht, pp. 181–199.

[167] McCrimmon, K.R. and Larsson, S. (1979). Utility theory: Axioms versus paradoxes, *in* M. Allais and O. Hagen (eds), *Expected utility hypotheses and the Allais paradox*, D. Reidel, pp. 27–145.

[168] McLean, J.E. and Lockwood, R.E. (1996). *Why and how should we assess students? The competing measures of student performance*, Sage Publications, Thousand Oaks.

[169] Merle, P. (1996). *L'évaluation des élèves. Enquête sur le jugement professoral*, PUF, Paris.

[170] Mintzberg, H., Raisinghani, D. and Théoret, A. (1976). The structure of unstructured decision processes, *Administrative Science Quarterly* **21**: 246–272.

[171] Mishan, E. (1982). *Cost-benefit analysis*, Allen and Unwin, London.

[172] Moom, T.M. (1997). *How do you know they know what they know? A handbook of helps for grading and evaluating student progress*, Grove Publishing, Westminster.

[173] Morisio, M. and Tsoukiàs, A. (1997). IUSWARE: A formal methodology for software evaluation and selection, *IEE Proceedings on Software Engineering* **144**: 162–174.

[174] Moscarola, J. (1984). Organizational decision processes and ORASA intervention, *in* R. Tomlinson and I. Kiss (eds), *Rethinking the process of operational research and systems analysis*, Pergamon Press, Oxford, pp. 169–186.

[175] Mousseau, V. (1993). *Problèmes liés à l'évaluation de l'importance en aide multicritère à la décision : Réflexions théoriques et expérimentations*, PhD thesis, LAMSADE, Université Paris-Dauphine, Paris.

[176] Munier, B. (1989). New models of decisions under uncertainty, *European Journal of Operational Research* **38**: 307–317.

[177] Nas, T.F. (1996). *Cost-benefit analysis: Theory and application*, Sage Publications, Thousand Oaks.

[178] Nauck, D. and Kruse, R. (1999). Neuro-fuzzy methods in fuzzy rule generation, *in* D. D. J. Bezdek and H. Prade (eds), *Fuzzy sets in approximate reasoning and information systems*, Vol. 3 of *Handbook of Fuzzy Sets*, Kluwer, Dordrecht, chapter 5, pp. 305–333.

[179] Nau, R.F. and McCardle, K.F. (1991). Arbitrage, rationality and equilibrium, *Theory and Decision* **31**: 199–240.

[180] Nau, R.F. (1995). Coherent decision analysis with inseparable probabilities and utilities, *Journal of Risk and Uncertainty* **10**: 71–91.

[181] Nguyen, H.T. and Sugeno, M. (1998). *Modelling and control*, Kluwer, Dordrecht.

[182] Nims, J.F. (1990). *Poems in translation: Sappho to Valéry*, The University of Arkansas Press, Arkansas.

[183] Noizet, G. and Caverini, J.-P. (1978). *La psychologie de l'évaluation scolaire*, PUF, Paris.

[184] Nurmi, H. (1987). *Comparing voting systems*, D. Reidel, Dordrecht.

[185] Nutt, P.C. (1984). Types of organizational decision processes, *Administrative Science Quarterly* **19**: 414–450.

[186] Nyborg, K. (1998). Some Norwegian politician's use of cost-benefit analysis, *Public Choice* **95**: 381–401.

[187] Ostanello, A. and Tsoukiàs, A. (1993). An explicative model of 'public' interorganizational interactions, *European Journal of Operational Research* **70**: 67–82.

[188] Ostanello, A. (1990). Action evaluation and action structuring – Different decision aid situations reviewed through two actual cases, *in* C.A. Bana

e Costa (ed.), *Readings in multiple criteria decision aid*, Springer Verlag, Berlin, pp. 36–57.

[189] Ostanello, A. (1997). Validation aspects of a prototype solution implementation to solve a complex MC problem, *in* J. Clímaco (ed.), *Multi-criteria analysis*, Springer Verlag, Berlin, pp. 61–74.

[190] Ott, W.R. (1978). *Environmental indices: Theory and practice*, Ann Arbor Science, Ann Arbor.

[191] Paschetta, E. and Tsoukiàs, A. (1999). A real world MCDA application: Evaluating software, *Technical report*, Document du LAMSADE N^o 113, Université Paris-Dauphine, Paris.

[192] Perny, P. and Pomerol, J.-Ch.. (1999). Use of artificial intelligence multicriteria decision making, *in* T. Gal, Th.J. Stewart and Th. Hanne (eds), *Advances in MCDM models, algorithms, theory, and applications*, Kluwer, Dordrecht, pp. 15.1–15.43.

[193] Perny, P. and Roubens, M. (1998). Fuzzy preference modelling, *in* R. Słowiński (ed.), *Fuzzy sets in decision analysis, operations research and statistics*, Kluwer, Dordrecht, pp. 3–30.

[194] Perny, P. and Zucker, J.D. (1999). Collaborative filtering methods based on fuzzy preference relations, *Proceedings of EUROFUSE-SIC'99*, pp. 279–285.

[195] Perny, P. (1992). Sur le non-respect de l'axiome d'indépendance dans les méthodes de type ELECTRE, *Cahiers du CERO* **34**: 211–232.

[196] Perrot, N., Trystram, G., Le Guennec, D. and Guely, F. (1996). Sensor fusion for real time quality evaluation of biscuit during baking. comparison between bayesian and fuzzy approaches, *Journal of Food Engineering* **29**: 301–315.

[197] Perrot, N. (1997). *Maîtrise des procédés alimentaires et théorie des ensembles flous*, PhD thesis, Ecole Nationale Supérieure des Industries Agricoles Alimentaires.

[198] Piéron, H. (1963). *Examens et docimologie*, PUF, Paris.

[199] Pirlot, M. and Vincke, Ph. (1997). *Semiorders. Properties, representations, applications*, Kluwer, Dordrecht.

[200] Pirlot, M. (1997). A common framework for describing some outranking procedures, *Journal of Multi-Criteria Decision Analysis* **6**: 86–93.

[201] Popham, W.J. (1981). *Modern educational measurement*, Prentice-Hall, New-York.

[202] Poulton, E.C. (1994). *Behavioral decision theory: A new approach*, Cambridge University Press, Cambridge.

[203] Quiggin, J. (1982). A theory of anticipated utility, *Journal of Economic Behaviour and Organization* **3**: 323–343.

[204] Quiggin, J. (1993). *Generalized expected utility theory – The rank-dependent model*, Kluwer, Dordrecht.

[205] Raiffa, H. (1970). *Decision analysis – Introductory lectures on choices under uncertainty*, Addison-Wesley, New York.

[206] Riley, H.J., Checca, R.C., Singer, T.S. and Worthington, D.F.. (1994). *Grades and grading practices: The results of the 1992 AACRAO survey*, American Association of Collegiate Registrars and Admissions Officers, Washington D.C.

[207] Rosenhead, J. (1989). *Rational analysis of a problematic world*, Wiley, New York.

[208] Roubens, M. and Vincke, Ph. (1985). *Preference modelling*, Springer Verlag, Berlin.

[209] Roy, B. and Bouyssou, D. (1991). Decision-aid: an elementary introduction with emphasis on multiple criteria, *Investigación Operativa* **2**: 95–110.

[210] Roy, B. and Bouyssou, D. (1993). *Aide multicritère à la décision : Méthodes et cas*, Economica, Paris.

[211] Roy, B. and Skalka, J.-M. (1984). ELECTRE IS : Aspects méthodologiques et guide d'utilisation, *Technical report*, Document du LAMSADE N^o 30, Université Paris-Dauphine, Paris.

[212] Roy, B. (1974). Critères multiples et modélisation des préférences : l'apport des relations de surclassement, *Revue d'Économie Politique* **1**: 1–44.

[213] Roy, B. (1990). Science de la décision ou science de l'aide à la décision ?, *Technical report*, Cahier du LAMSADE N^o 97, Université Paris-Dauphine, Paris.

[214] Roy, B. (1993). Decision science or decision-aid science?, *European Journal of Operational Research* **66**: 184–204.

[215] Roy, B. (1996). *Multicriteria methodology for decision aiding*, Kluwer, Dordrecht. Original version in French "*Méthodologie multicritère d'aide à la décision*", Economica, Paris, 1985.

[216] Russo, J.E. and Schoemaker, P.J.H. (1989). *Confident decision making*, Piatkus, London.

[217] Saaty, T.L. (1980). *The analytic hierarchy process*, McGraw-Hill, New York.

[218] Sabot, R. and Wakeman, L.J. (1991). Grade inflation and course choice, *Journal of Economic Perspectives* **5**: 159–170.

[219] Sager, C. (1994). *Eliminating grades in schools: An allegory for change*, A S Q Quality Press, Milwaukee.

[220] Salles, M., Barrett, C.R. and Pattanaik, P.K. (1992). Rationality and aggregation of preferences in an ordinally fuzzy framework, *Fuzzy Sets and Systems* **49**: 9–13.

[221] Satterthwaite, M.A. (1975). Strategy proofness and Arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions, *Journal of Economic Theory* **10**: 187–217.

[222] Savage, L. (1954). *The foundations of statistics*, 1972, 2nd revised edn, Wiley, New York.

[223] Schmeidler, D. (1989). Subjective probability and expected utility without additivity, *Econometrica* **57**: 571–587.

[224] Schneider, Th., Schieber, C., Eeckoudt, L. and Gollier, C. (1997). Economics of radiation protection: Equity considerations, *Theory and Decision* **43**: 241–51.

[225] Schofield, J. (1989). *Cost-benefit analysis in urban and regional planning*, Unwin and Hyman, London.

[226] Scotchmer, S. (1985). Hedonic prices and cost-benefit analysis, *Journal of Economic Theory* **37**: 55–75.

[227] Sen, A.K. (1986). Social choice theory, *in* K.J. Arrow and M.D. Intriligator (eds), *Handbook of mathematical economics*, Vol. 3, North-Holland, Amsterdam, pp. 1073–1181.

[228] Sen, A.K. (1997). Maximization and the act of choice, *Econometrica* **65**: 745–779.

[229] Simon, H.A. (1957). *A behavioural model of rational choice* in *Models of man*, Wiley, New York, pp. 241–260.

[230] Sinn, H.W. (1983). *Economic decisions under uncertainty*, North-Holland, Amsterdam.

[231] Słowiński, R. (ed.) (1998). *Fuzzy sets in decision analysis, operations research and statistics*, Kluwer, Dordrecht.

[232] Sopher, B. and Gigliotti, G. (1993). A test of generalized expected utility theory, *Theory and Decision* **35**: 75–106.

[233] Speck, B.W. (1998). *Grading student writing: An annotated bibliography*, Greenwood Publishing Group, Westport.

[234] Stamelos, I. and Tsoukiàs, A. (1998). Software evaluation problem situations, *Technical report*, Cahier du LAMSADE N^o 156, Université Paris-Dauphine, Paris.

[235] Steuer, R.E. (1986). *Multiple criteria optimisation: Theory, computation, and application*, Wiley, New York.

[236] Stratton, R.W., Myers, S.C. and King, R.H. (1994). Faculty behavior, grades and student evaluations, *Journal of Economic Education* **25**: 5–15.

[237] Sugden, R. and Wiliams, A. (1983). *The principles of practical cost-benefit analysis*, Oxford University Press, Oxford.

[238] Sugeno, M. (1977). Fuzzy measures and fuzzy integrals: a survey, *in* M.M. Gupta, G.N. Saridis and B.R. Gains (eds), *Fuzzy automata and decision processes*, North Holland, Amsterdam, pp. 89–102.

[239] Sugeno, M. (1985). An introductory survey on fuzzy control, *Information Sciences* **36**: 59–83.

[240] Suzumura, K. (1999). Consequences, opportunities and procedures, *Social Choice and Welfare* **16**: 17–40.

[241] Syndicat des Transports Parisiens (1998). Méthodes d'évaluation des projets d'infrastructures de transports collectifs en région Ile-de-France, *Technical report*, Syndicat des Transports Parisiens, Paris.

[242] Tchudi, S. (1997). *Alternatives to grading student writing*, National Council of Teachers of English, Urbana.

[243] Teghem, J. (1996). *Programmation linéaire*, Éditions de l'Université de Bruxelles-Éditions Ellipses, Brussels.

[244] Thaler, R.H. (1991). *Quasi rational economics*, Russell Sage Foundation, New York.

[245] Toth, F.L. (1997). *Cost-benefit analysis of climate change: The broader perspectives*, Birkhäuser, Basel.

[246] Trystram, G., Perrot, N. and Guely, F. (1995). Application of fuzzy logic for the control of food processes, *Processing Automation* **4**: 504–512.

[247] Tsoukiàs, A. and Vincke, Ph. (1995). A new axiomatic foundation of partial comparability, *Theory and Decision* **39**: 79–114.

[248] Tsoukiàs, A. and Vincke, Ph. (1999). A characterization of PQI interval orders, *Proceedings OSDA '98*, Electronic Notes on Discrete Mathematics, pp. (`http://www.elsevier.nl/locate/endm`), to appear also in Discrete Applied Mathematics.

[249] Tversky, A. (1969). Intransitivity of preferences, *Psychological Review* **76**: 31–48.

[250] United Nations Development Programme (1997). *Human Development Report 1997*, Oxford University Press, Oxford.

[251] van Doren, M. (1928). *An anthology of world poetry*, Albert and Charles Boni, New York.

[252] Vansnick, J.-C. (1986). De Borda et Condorcet à l'agrégation multicritère, *Ricerca Operativa* (40): 7–44.

[253] Vassiloglou, M. and French, S. (1982). Arrow's theorem and examination assessment, *British Journal of Mathematical and Statistical Psychology* **35**: 183–192.

[254] Vassiloglou, M. (1984). Some multi-attribute models in examination assessment, *British Journal of Mathematical and Statistical Psychology* **37**: 216–233.

[255] Vincke, Ph. (1988). P, Q, I preference structures, *in* J. Kacprzyk and M. Roubens (eds), *Non conventional preference relations in decision making*, Springer Verlag, Berlin, pp. 72–81.

[256] Vincke, Ph. (1992a). Exploitation of a crisp binary relation in a ranking problem, *Theory and Decision* **32**: 221–241.

[257] Vincke, Ph. (1992b). *Multi-criteria decision aid*, Wiley, New York. Original version in French "*L'Aide Multicritère à la Décision*", Éditions de l'Université de Bruxelles-Éditions Ellipses, Brussels, 1989.

[258] Viscusi, W.K. (1992). *Fatal tradeoffs: Public and private responsibilities for risk*, Oxford University Press, Oxford.

[259] von Neumann, J. and Morgenstern, O. (1944). *Theory of games and economic behavior*, Princeton University Press, Princeton.

[260] von Winterfeldt, D. and Edwards, W. (1986). *Decision analysis and behavioral research*, Cambridge University Press, Cambridge.

[261] Wakker, P.P. (1989). *Additive representations of preferences – A new foundation of decision analysis*, Kluwer, Dordrecht.

[262] Warusfel, A. (1961). *Les nombres et leurs mystères*, Points Sciences, Seuil, Paris.

[263] Watson, S.R. (1981). Decision analysis as a replacement for cost-benefit analysis, *European Journal of Operational Research* **7**: 242–248.

[264] Weinstein, M.C. and Stason, W.B. (1977). Foundations of cost-effectiveness analysis for health and medical practices, *New England Journal of Medicine* **296**: 716–721.

[265] Weitzman, M.L. (1994). On the "environmental" discount rate, *Journal of Environmental Economics and Management* **26**: 200–209.

[266] Weymark, J.A. (1981). Generalized Gini inequality indices, *Mathematical Social Sciences* **1**: 409–430.

[267] Willis, K.G., Garrod, G.D. and Harvey, D.R. (1998). A review of cost-benefit analysis as applied to the evaluation of new road proposals in the U.K., *Transportation Research – D* **3**: 141–156.

[268] Yaari, M.E. (1987). The dual theory of choice under risk, *Econometrica* **55**: 95–115.

[269] Yu, W. (1992). *Aide multicritère à la décision dans le cadre de la problématique du tri : Méthodes et applications*, PhD thesis, LAMSADE, Université Paris-Dauphine, Paris.

[270] Zadeh, L.A. (1979). A theory of approximate reasoning, *in* J.E. Hayes, D. Michie and L.I. Mikulich (eds), *Machine intelligence*, Elsevier, Amsterdam, pp. 149–194.

[271] Zadeh, L.A. (1999). From computing with numbers to computing with words. from manipulation of measurement to manipulation of perceptions, *Proceedings of EUROFUSE-SIC'99*, pp. 1–2.

[272] Zarnowsky, F. (1989). *The decathlon – A colorful history of track and field's most challenging event*, Leisure Press, Champaign.

[273] Zerbe, R.O. and Dively, D.D. (1994). *Benefit-cost analysis in theory and practice*, Harper Collins, New York.

Index

absolute scale, 115
action, 212
actor, 206
acyclic, 126
aggregation, 30, 41, 51, 148, 245
  weighted sum, 172
  additive, 44
  compensation, 46, 57, 212
  conjunctive rule, 41, 96
  constructive approach, 130
  disaggregation, 117
  dominance, 93
  linearity, 46, 80
  monotonicity, 10, 61
  multi-attribute value function, 105
  non-compensation, 141
  paired comparison, 193
  procedure, 215
  rank reversal, 117
  screening process, 96
  single-attribute value function, 106
  tournament, 125
  utility, 105
  utility function, 106
  value function, 106
  weight, 35
  weighted average, 42, 59, 85, 241
  weighted sum, 155, 159, 166
AHP, 111
  rank reversal, 117
air quality, 61, 63
Allais' paradox, 191
ambiguity, 212
analyst, 206
anchoring effect, 34
Arrow, 16
aspiration level, 96
astrology, 237
attributes
  hierarchy, 214
attributes hierarchy, 213
automatic decision, 239
automatic decision systems, 148
axiomatic analysis, 244

bayesian decision theory, 239, 244
binary relation
  acyclic, 126
  fuzzy, 21
  incomparability, 19, 130, 220
  outranking, 105
  semiorder, 20
  transitivity, 18–20
Borda, 14
Borda's method, 124

call for tenders, 208
cardinal, 47
client, 206
coalition, 219
coherence test, 214
communication, 242
compensation, 46, 57, 212
computer science, 30, 237
concordance, 216, 219
concordance threshold, 134
Condorcet, 13
  paradox, 51
Condorcet's method, 125
conjunctive rule, 41, 96
consistency, 84
constructive approach, 130
corporate finance, 73
correlation, 102
cost-benefit analysis, 71, 238

externalities, 78
markets, 76
net present social value, 76
price, 75, 85
price of human life, 81
price of time, 80
public goods, 78
social benefits, 75
social costs, 75
social welfare, 77, 86
credibility index, 139
criteria
coalition, 219
coherence test, 214
coherent family, 214
hierarchy, 213, 214
interaction, 103
point of view, 212
relative importance, 216, 219
cycle reduction, 135

decathlon, 63, 66
decision
dynamic consistency, 202, 239
legitimation, 220
model, 1
decision aiding process, 206
decision model, formal, 84, 212, 237
decision process, 85, 206
actor, 206
analyst, 206
client, 206
decision rule, 149, 151
decision support, 210
evaluation model, 206, 213
final recommendation, 219
learning process, 119
problem formulation, 206, 211, 212
problem situation, 206
problem statement, 212, 215
decision table, 152
decision theory, 237
Allais' paradox, 191
Ellsberg's paradox, 192
expected utility, 189, 201
expected value, 187
St. Petersburg game, 187
disaggregation, 117
discordance, 138
discounting, 74, 86, 239
net present value, 74
social rate, 76, 82
dominance, 51, 93, 96
dynamic consistency, 202, 239

economics, 71, 237
education science, 237
elections, 237
ELECTRE-TRI, 215, 228
Ellsberg's paradox, 192
engineering, 30, 237
environment, 71, 82
equity, 79, 86
evaluation
absolute, 212
model, 1, 40, 51, 206, 213
problem statement, 215
software, 207, 213
evaluation model
problem statement, 212
expected value, 187
externalities, 78

final recommendation, 219
forecasting, 80
fuzzy, 21
control, 149, 165
implication, 166
interval, 161
labels, 161
rule, 169
set, 161, 169

GPA, 48
grade, 29, 237
anchoring effect, 34
GPA, 48
marking scale, 33
minimal passing, 36, 42
standardised score, 34
graphology, 237

health, 71
heuristics, 242
hierarchy, 213, 214
human development, 54, 61

ideal point, 117
implication, 166
imprecision, 51, 103
incomparability, 19, 130, 220
independence, 58, 102
 of irrelevant alternatives, 15
 separability, 214
indicator, 238
indices, 173
indifference threshold, 201
interaction, 103
interactive methods, 105
interpolation, 155
interval scale, 99
intuition, 242

kernel, 131

learning process, 119
legitimation, 220
linear scale, 104

majority rule, 125
manipulability, 17
markets, 76
marking scale, 33
mathematics, 30
MCDM, 238
 ideal point, 117
 interactive methods, 105
 sorting, 212
  profiles, 220
 substitution rate, 57
 swing-weight, 110
 trade-off, 101
meaningfulness, 62, 227
measurement, 38, 51, 67, 212
 absolute scale, 115
 cardinal, 47
 interval scale, 99
 linear scale, 104
 meaningfulness, 62, 227
 nominal scale, 214
 ordinal, 39, 47, 214
 ratio scale, 98
 reliability, 33
 scale, 32, 38, 79
 standard sequences, 107
 subjective, 214
 validity, 33
model, 40, 245
 structuration, 84, 242, 245
mono-criterion analysis, 76, 85
monotonicity, 10, 61

nearest neighbours, 172
net present social value, 76
net present value, 74
nominal scale, 214
non-compensation, 141

operational research, 30, 237
ordinal, 39, 47, 214
outranking, 105
outranking methods, 124, 129
 discordance, 138
 concordance, 216, 219
 concordance threshold, 134
 credibility index, 139
 cycle reduction, 135
 ELECTRE-TRI, 215, 228
 incomparability, 130
 indifference threshold, 201
 majority rule, 125
 PROMETHEE, 193
 veto, 216, 219

paired comparison, 193
point of view, 212
political science, 237
preference
 model, 214, 245
 nontransitive, 130
 relation, 125
 threshold, 50
price, 75, 85
price of human life, 81
price of time, 80
priority, 111

probability, 239
problem formulation, 206, 211, 212
problem situation, 206
problem statement, 212, 215
PROMETHEE, 193
public goods, 78

rank reversal, 117
ranking, 212
ratio scale, 98
relative importance, 216, 219
risk, 239
robustness, 86, 242, 245
rule
    aggregation, 148

scale, 32, 38, 79
screening process, 96
security, 81
semiorder, 20
sensitivity analysis, 83
    stability, 99, 101
separability, 214
similarity
    indices, 173
    relation, 173
social benefits, 75
social costs, 75
social rate, 76, 82
social welfare, 77, 86
software, 207, 213
sorting, 212
St. Petersburg game, 187
stability, 99, 101
statistics, 237
structuration, 84, 242, 245
subjective, 214
substitution rate, 57

t-norm, 164
threshold, 50, 173
tournament, 125
trade-off, 101
transitivity, 18–20
transportation, 71, 79

unanimity, 13
uncertainty, 51, 79, 103, 179, 182, 201, 239
    endogenous, 218, 221
    exogenous, 218
utility, 105, 106
    expected, 189, 201

value function, 106
    multi-attribute, 105
    single-attribute, 106
veto, 216, 219
voting procedure
    Borda's method, 124
    Concordet paradox, 51
    Condorcet's method, 125
    manipulability, 17
    unanimity, 13

weight, 35
weighted average, 42, 59, 85, 241
weighted sum, 155, 159, 166, 172