

# Multiattribute value elicitation

**Alec Morton**

University of Strathclyde

## **Abstract**

Multiattribute Value Theory (MAVT) methods are perhaps the most intuitive multicriteria methods, and have the most theoretically well-understood basis. They employ a divide-and-conquer modelling strategy in which the value of an option is conceptualised as a function (typically the sum) of the scores associated with the performance of the option on different attributes. This chapter outlines the concept of preferential independence, which has a critical underpinning role of elicitation within the MAVT paradigm. I also present MAVT elicitation in the context of the overall Decision Analysis process, comprising three broad stages: Designing and Planning; Structuring the Model; and Analysing the Model. I outline some of the main practical methods for arriving at the partial values and weighting them to arrive at an overall value score, including both traditional methods relying on cardinal assessment, and the MACBETH approach which uses qualitative difference judgements. A running example of a house choice problem is used to illustrate the different elicitation approaches.

## **Background**

The Multi-Attribute Value Theory (MAVT) approach, and in particular the additive model, is perhaps the most intuitive of all Multi-Criteria Decision Analysis (MCDA) methods. The decision aiding procedure suggested by MAVT is to line up the options, compare them according to a common set of criteria, assign scores to each option according to their performance on each criterion, weight these criteria and calculate an overall score for each option. The computations involved in applying MAVT are relatively straightforward compared to the methods of the outranking school (see Chapter 14), and hence the method is transparent and easily understood. One of the insights of this MAVT paradigm is that this seemingly simple procedure, involving nothing more than elementary arithmetic, actually re-

quires quite a high level of conceptual sophistication to use well and appropriately.

The need for conceptual sophistication arises when one attempts to specify formally the meaning of the scores and weights in the procedure of the previous paragraph. The meaning of a probability, by contrast, is relatively clear, in the following sense. Although the exact interpretation of a probability statement depends on one's preferred axiomatics (French, 1986; French and Rios Insua, 2000), probabilities are ultimately rooted in the procedure of counting which is a natural first step on the path to quantification. If an assessor is well-calibrated, of the class of events she assesses as having probability 50%, half will be realized, and half will not.

By contrast, value is not rooted in counting, but in preferring. However, whereas counting establishes an association between a set of things and a number, preferring merely establishes a relationship between two things: one thing is better than, more attractive than, or more desirable than, another. From such a binary relation, it is easy to see how to establish a ranking of objects. However, how might one go about associating *numbers* to options according to their criterion-wise performance in a principled way?

The central concept of MAVT is that as well as possessing an idea of *preference*, we also possess an idea of *strength of preference* (Dyer and Sarin, 1979; Köbberling, 2006). Thus, when thirsty on a hot day, I may have a slight preference for iced tea over iced coffee, but a strong preference for an iced drink over no iced drink. The difference between the scores I give to iced tea and iced coffee should therefore be relatively small, but the difference between these scores and no iced drink should be relatively large. However, unlike preferences, which can be observed by an outside party who studies the elicitee's choice behavior (I offer you a menu consisting of iced tea and coffee and see which, if either, you choose), strengths of preference are not observable. Nevertheless, the concept seems to be one which is intuitive and natural to most of us from casual introspection and ordinary discourse.

An alternative way to assign numbers to multiattributed options is the Multi-Attribute Utility Theory (MAUT) approach (dealt with in Chapter 9). MAUT, like MAVT, provides a framework for deriving scores and weights. However, the interpretation of the scores and weights in MAUT does not use a strength of preference concept – rather it uses an approach based on equivalent gambles. MAUT is necessary if we are dealing with uncertain events, for instance in a multiattribute decision tree. However, while the MAUT mode of questioning can be appropriate

in many settings, it presupposes a facility with probabilistic thinking which many people do not have, and involves asking questions which many people find confusing and irrelevant.

In this chapter, I do the following. I begin with a discussion of the concept of preferential independence which is a foundational concept in the use of scoring and weighting methods based on MAVT. The main section presents MAVT elicitation in the context of the decision analysis process, from establishing aims through to sensitivity analysis and stress testing of the model. To assist readers who may be interested in using these procedures, I also provide some “troubleshooting” hints and tips. I conclude with some suggestions for future prospects for MAVT methods. The interested reader is referred for comparison to other textbooks which deal with similar material such as Goodwin and Wright (2014) and Howard and Abbas (2016) as well as the seminal text of von Winterfeldt and Edwards (1986).

## **Preferential independence: a foundational concept of Multiattribute Value Theory**

A natural starting point is to ask the question: under what circumstances can MAVT be used? As it happens there is a very clear and mathematically well-specified answer to this question (Krantz et al, 1971; French, 1986). To explain this answer I introduce the idea of a representation theorem. Representation theorems connect qualitative properties of preferences with functions which represent these preferences. (A function is said to represent preferences if it assigns a higher number to a more preferred object). Representation theorems have the following generic two-part form. The main action revolves around the relation  $\succeq$ , read “is weakly preferred to” or “is at least as good as”.

### **Generic Representation Theorem**

1. (Sufficiency) If the relation  $\succeq$  has such and such properties, then there exists a real valued function  $v(\bullet)$  of such and such a form such that  $a \succeq b$  if and only if  $v(a) \geq v(b)$ .
2. (Necessity) If there exists a real valued function  $v(\bullet)$  of such and such a form such that  $a \succeq b$  if and only if  $v(a) \geq v(b)$ , then the relation  $\succeq$  has such and such properties.

Note the differing role of these two parts of the theorem: the sufficiency part tells us that if an elicitee has a preference relation with certain characteristics, then there exists a real value function, whereas the necessity part tells us the opposite. (In general the sufficiency part is harder to prove than the necessity part.)

An example of a representation theorem (Krantz et al., 1971) is the following theorem which guarantees the existence of a general function.

**Representation Theorem for the existence of a representing function.**

1. (Sufficiency). If  $\succsim$  is complete and transitive, then there exists a  $v(\bullet)$  which represents  $\succsim$ .
2. (Necessity). If there exists a  $v(\bullet)$  which represents  $\succsim$ , then  $\succsim$  is complete and transitive.

What this tells us is that if an elicitee has preferences which are non-transitive – she tells us she prefers tea to coffee and coffee to hot chocolate, and hot chocolate to tea, there is no representing function for her preferences. A moment's reflection shows why this is so: it would require finding three numbers  $x$ ,  $y$  and  $z$  such that  $x > y$ ,  $y > z$ , and  $z > x$ , which is plainly impossible.

A more interesting and subtle question is under what circumstances can scoring and weighting can be used to arrive at an evaluation of options. Scoring and weighting implicitly involves the use of an additive value function

$$v(a) = \sum_j w_j v_j(a)$$

, where  $v_j$  is a scoring function which assigns scores for

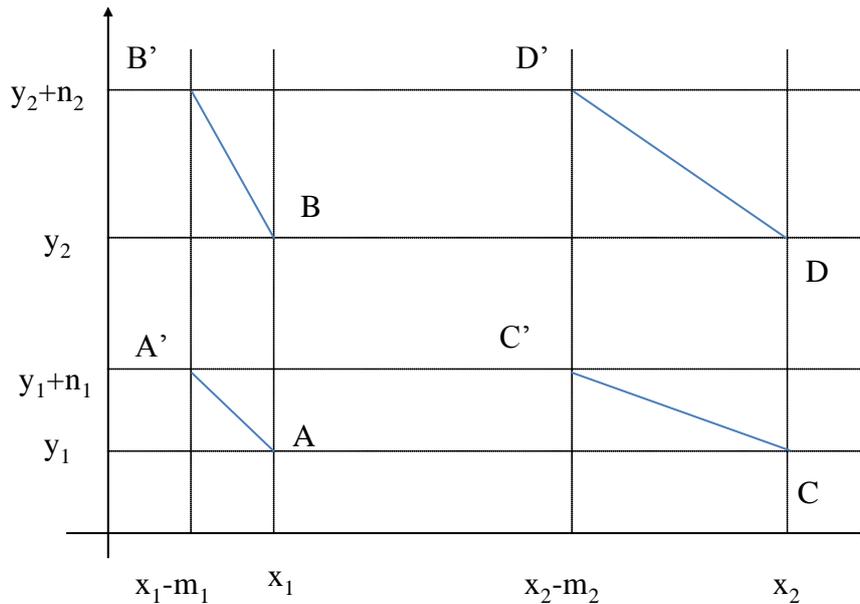
each criterion  $j$  to each option  $a$  and  $w_j$  is the weight of criterion  $j$ . Is there a representation theorem which tells us when this value function can be used? As it happens, there are several such representation functions. One useful illustrative example is the following.

**Representation theorem for the existence of an additive representing function**

Let  $\succsim$  be a preference ordering on a set of biattributed options with well-defined partial preferences  $\succsim_i$  for  $i=1$  and  $2$ . Given certain technical assumptions, the Reidemeister condition is necessary and sufficient for the existence of a representing additive value function.

As this chapter aims for informality, I do not propose to explain this Theorem in detail here. In particular I ignore the role of technical conditions such as solvabil-

ity and the Archimedean axiom in proving the result. However, the Reidemeister condition is insightful and it is worth taking some time to present in detail. To understand the condition, consider Figure 1 (the illustration is based on that in Belton and Stewart, 2002).



**Fig. 1.** Four pairs of points in a biattribute space, illustrating the Reidemeister condition.

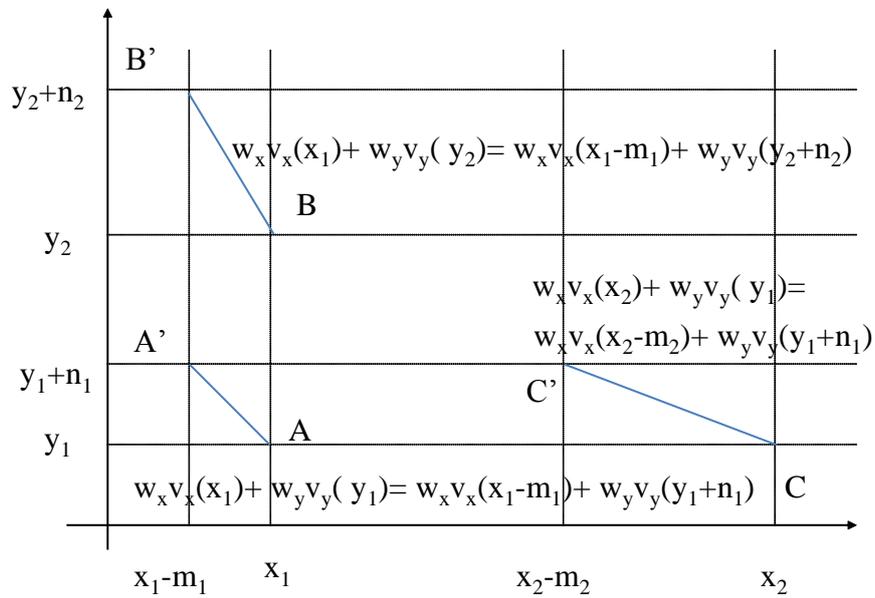
Figure 1 shows points in a biattribute space, with dimensions  $x$  and  $y$ . For example, in choosing a house,  $x$  and  $y$  could be *square footage* and (the negative of) *purchase price*.  $A$  and  $A'$ ;  $B$  and  $B'$  and  $C$  and  $C'$  are pairs of points in this space (each pair representing a larger, more expensive house, and a smaller, cheaper house) between which the elicitee is indifferent, i.e. prefers neither one nor the other.

The Reidemeister condition is a condition on the elicitee's preferences. An elicitee's preferences obey this condition if, whenever she is indifferent between  $A$  and  $A'$ ;  $B$  and  $B'$  and  $C$  and  $C'$  respectively, she is also indifferent between  $D$  and  $D'$ . To see why this condition is sufficient for the existence of a representing additive value function is hard: the proof involves using the condition iteratively to construct a grid of points which have the interpretation of a value function of the additive form. But to see the necessity is easy. Consider Figure 2. If the elicitee's preferences are represented by an additive value function, then the formulae for

the values of these indifferences can be written as shown on the grid. The reader can verify that by adding the equations corresponding to the indifferences between B and B' and C and C' respectively, and subtracting the equation corresponding to A and A', the result is the following:

$$w_x v_x(x_2) + w_y v_y(y_2) = w_x v_x(x_2 - m_2) + w_y v_y(y_2 + n_2)$$

But this equation expresses nothing other than the idea that D is indifferent to D'. Hence any elicitee whose preferences are represented by an additive function must obey the Reidemeister condition.

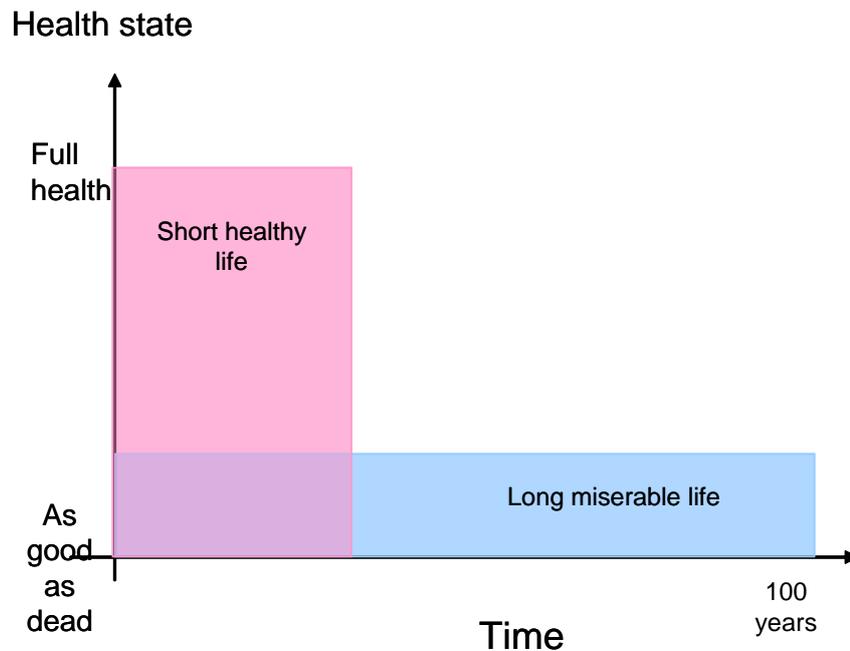


**Fig. 2.** Value functions associated with the indifferences between A and A', B and B', and C and C'.

It is not always or necessarily the case that the Reidemeister condition holds. In the case of buying a house, I may feel that the value of difference in space  $m_2$  depends on the price which I am prepared to pay for the house. When I pay a lower price for the house, I can use the space to host fabulous parties, and hence the additional space has some value to me. But when I pay a higher price, I have no spare money for entertaining and the additional space just means that I have to

spend more time cleaning. Hence, it does not make sense to give “points” to the additional space irrespective of the financial purchase price of the house.

If the Reidemeister condition or its equivalents fail to hold, that does not necessarily mean that all is lost. There are models which represent situations where there are interactions between criteria. The simplest and most intuitive example is that the “Quality Adjusted Life Year”, or QALY, which has found widespread use in health economics as a measure of health benefit associated with a life extension or enhancement (for axiomatics, see Pliskin et al, 1980; Miyamoto et al, 1998). At its simplest the key idea of the QALY is that an individual’s life can be considered as characterised in two dimensions: length and quality of life. Figure 3 illustrates two individuals, one of whom enjoys a short healthy life and the other of whom experiences a long miserable life.



**Fig. 3. Two possible lifecourses**

For health gains, it makes no sense to calculate the value of a health gain as a weighted sum of duration and quality of life. To see why not, consider the extreme case of a life extension of zero (or infinitesimal) duration. Such a life extension clearly has no value, no matter how good the health state. This is not compatible with an additive model where the contribution of a set number of years

of life to overall value is fixed, independently of the number of years lived in that health state. For this reason, QALYs are calculated as the length of life *multiplied* by a factor representing the quality of life (this can be visualised as the area of the rectangles in Figure 3.). Indeed, one popular way to elicit the value of a health state is to ask a so-called time tradeoff question, where a number is associated with a health state  $h$  (being blind, for example) by asking the elicitee for a number of years  $n$  such that they would be indifferent between  $n$  years in state  $h$  and one year in full health (Drummond et al, 2015).

### The decision analysis process

Having sketched the foundational concept of preferential independence, I now turn to the question of how to actually elicit scores and weights. Attempting to elicit scores and weights in the context of a poorly specified decision problem is a hopeless undertaking: before elicitation can take place, the problem context, and the basic elements of the model must be clearly specified and understood by all relevant parties in the elicitation. Accordingly I will structure this chapter through a map of the decision analysis process (see Figure 4.).

<b><i>Design and planning</i></b>	
Step 1.	Establish the aims of the analysis
Step 2.	Identify decision makers, stakeholders, and persons with relevant expertise
Step 3.	Design the intervention
<b><i>Structuring the model</i></b>	
Step 4.	Identify the options
Step 5.	Identify the criteria
Step 6.	Score the options on the criteria
Step 7.	Weight the criteria
<b><i>Analysing the model</i></b>	
Step 8.	Compute overall ranking
Step 9.	Conduct sensitivity analysis

**Fig 4. Schematic of the decision analysis process.**

## ***Design and planning***

### **Step 1. Establish the aims of the analysis**

A sensible starting point is to identify the objectives of the decision. For example, the objectives could be: to grow the organization (in terms of revenues, reputation, market share, or profitability); to contribute to social welfare (e.g. through the provision of healthcare or recreation facilities); to contribute to equity objectives (for example health equity, equity in income distribution); to contribute to some other stated policy objective (such as reducing error in tax collection or benefits payment); or to help an organisation (e.g. a government agency or social enterprise) fulfil its mission.

The analysis may be intended to support different problem statements or *problématiques* (Roy, 1985):

- Single choice (choose one option from  $n$  options) – for example choosing a site for a new airport.
- Multiple choice (choose  $k$  options from  $n$  options) – for example members of a team or a board.
- Budget allocation (choose options subject to a budget constraint of  $B$ ) – for example determining a portfolio of R&D projects, or military equipment for purchase.
- Development of a priority ordering – for example ranking applicants for a scholarship in terms of their merit.
- Accepting or rejecting an option (for example, deciding whether a new drug can be provided by the national healthcare system).

Articulating both aims and the *problématique* is often a useful starting point for analysis.

### **Step 2. Identify decision makers, stakeholders, and persons with relevant expertise**

It is important to identify early on both the decision makers, stakeholders, who may be individuals, organisational units, or organisations, and persons with relevant expertise. A decision maker is someone who has the authority to make a decision. A common definition of a decision is that it is “an irrevocable allocation of resources, in the sense that it would take additional resources, perhaps prohibitive in amount, to change the allocation” (Matheson and Howard, 1983). Thus, to qualify as a decision maker, one must have the power to allocate resources. A

stakeholder is someone [who](#) can affect or is affected by a decision (for interesting discussions of the stakeholder concept, see Bryson, 2004; Ackermann and Eden, 2011). An expert, by contrast, is someone who has knowledge relevant to the assessment of the characteristics of the options at hand (for more discussion, see European Food Safety Authority, 2014, Appendix A.2.2).

### Step 3. Design the intervention

Often, MAVT is used in a participative way – in what Montibeller and Franco (2010) call the “facilitated mode” of analysis. Sometimes, the entire decision analysis process will take place in a workshop or series of workshops (this is sometimes known as “decision conferencing” – Phillips, 2007). Workshops are often valuable as they build consensus and enable disagreements to be explored and sometimes resolved, however, they can be time-consuming and expensive. On other occasions, analysis may be done entirely “in the backroom” – such behind the scenes analysis can still be valuable contribution to clarifying the problem and guiding a path to a decision.

Sometimes, it may be most useful to have a hybrid process. For example, scoring can be done “offline” by individuals, so that when face-to-face discussion takes place it can focus on where there are differences of opinion in the scoring. In thinking through the design of an intervention, it may be useful to fill in a matrix of the form shown in Table 1.

	Who to involve?	How to involve?
Options		
Criteria		
Scoring		
Weighting		
Sensitivity analysis		

Table 1. Matrix for determining involvement in a MAVT application

Different modes of working may make sense in different contexts. For example, when options are scientific projects which contribute to public welfare, it may make sense to have scientists identify the options and perform the scoring on an individual basis, but for representatives of the relevant stakeholders to do weighting in workshop.

## ***Structuring the model***

### **Step 4. Identify the options**

Options (sometimes called alternatives or actions) are things which could be done.

Options should be:

- Creative. It is important to canvass a wide range of options, even options which are not immediately doable.
- Manageable in the time available. The number of options drives the length of time required by the analysis.
- Homogeneous – they should be the same sort of thing. For example
  - a facility which will deliver benefits over a 5 year timeframe cannot be directly compared with a facility which will deliver benefits over a 100 year timeframe.
  - an investment option which costs £50 cannot be directly compared with an investment option which costs £1,000,000.
- If more than one option can be done, options should be evaluatively independent, that is, it should be possible to evaluate an option *a* without knowing whether a second option *b* is to be implemented.

Let us look at an example where evaluative independence might fail. I cannot evaluate “coffee” without knowing whether I am also to receive “milk” (as it happens, I prefer not to drink my coffee black) and vice versa. If options are not evaluatively independent they can sometimes be restructured to achieve evaluative independence (e.g. I combine “coffee” and “milk” into a single option). Sometimes this is not possible, and more complex approaches are required, such as the use of mathematical programming methods.

It is good practice to identify a baseline level of activity (“do nothing”). This is particularly important where the problem is not a problem of single choice. In multiple choice contexts, failure to identify an appropriate baseline can lead to paradoxical behavior where model results change depending on seemingly arbitrary features of model specification – see Morton (2015) for more details.

### **Step 5. Identify the criteria**

Criteria are the measures of performance by which an option is judged. Just be-

cause in MAVT - and indeed in Multi-Criteria Decision Analysis (MCDA) procedures more generally – the aim is to identify criteria which can be used to guide choice, this does not mean that these criteria really “exist” in the world: they have to be discussed, negotiated and agreed between the various decision makers. Indeed, research tells us that people are often not even sure what their own objectives are, even in problems which are quite important to them (Bond et al, 2010): this is a reason why there needs to be a structured process to discuss these objectives and arrive at a model which everyone can sign up to.

A useful question to identify criteria is consider the options and ask the question “what would distinguish between a good and bad choice in this decision problem?” Criteria thus form a bridge between the options and the objectives.

Criteria have a sense or direction of preference:

- if one prefers more of the criterion to less (e.g. revenue), one says it has an *increasing direction of preference*
- if one prefers less to more (e.g. cost) one says it has a *decreasing direction of preference*.

Once criteria have been identified, it should be possible to describe how the options perform against the criteria. This can be done by specifying a performance matrix, with options along the vertical dimension and criteria along the horizontal direction. The individual performances are described in the cells: these can be described either in terms of natural attributes (e.g. number of lives saved); constructed attributes (e.g. numbers of stars which summarise further disaggregate information); or qualitative descriptions (e.g. “very good”; “barely adequate”).

Suppose one is choosing a house to purchase. Table 2 shows an example of a performance matrix (this example also appears in Morton and Fasolo, 2009). Here, Financial Cost is operationalized through money (in £); Closeness to the city centre is operationalized through the zone of the city in which the house is located (A is closest to the centre and C is furthest way); Character is assessed as a simple “yes” or “no”; and size is measured in square footage. The measures which are used to operationalize the criteria are called *attributes*: unlike criteria which are expressions of a decision maker’s aspirations in a decision problem, attributes are objective characteristics which can be “read off” from a description of the options themselves.

	Criteria			
House	Financial Cost (£)	Closeness (Zone)	Character	Size (Sq footage)
1	220	A	Yes	600
2	180	B	Yes	600
3	130	C	No	700
4	120	C	No	500
5	180	B	No	600

Table 2. Performance matrix for the house choice problem

Sometimes it is possible to identify options which are *dominated*. An option  $a$  is said to be dominated by a second option  $b$  if  $b$  is at least as good as  $a$  on each criterion and strictly better than  $a$  on at least one criterion. In single choice problems, dominated options will always be ranked at least second, and so can be eliminated from consideration. For example, in the house choice problem, House 5 is dominated by House 2. It performs the same as House 2 on every criterion except Character: House 2 has character and House 5 has no character.

If there are a large number of criteria, it may be worthwhile structuring the criteria as a hierarchical value tree (see e.g. Figure 2 of Chapter 9). As a whole, the set of criteria should be (Keeney and Raiffa, 1976):

- *Discriminatory*. They should distinguish between options. Sometimes there may be objectives which are felt to be very important but which do not distinguish between the options under consideration (e.g. how a software program is designed may have no impact on climate change). In this case, there will be no criterion associated with this objective in this decision problem.
- *Complete*. Criteria should capture everything which the decision makers and stakeholders care about.
- *Small in number*. As with options, a large number of criteria result in options will increase time and care should be taken not to list too many criteria.
- *Non-redundant*. Criteria should not duplicate each other: there should be no double counting.

- *Preferentially Independent* (as discussed earlier in this chapter). One useful way to test whether preferential independence holds in practice is to see whether it is possible for the elicitee to assess the value of performance on one criterion independently of the level of performance on another criterion. If not, this suggests that preferential independence does not apply and so the model should be restructured, or a non-additive value model should be applied.

As I have stressed above, preferential independence is critical if scoring and weighting approaches are to be used. Here is an example where preferential independence might fail in our house choice setting. In choosing a house a purchaser may care about whether there is a park nearby, and about whether there is a swimming pool nearby: but if there is a park, she no longer care so much about the swimming pool (and *vice versa*). Often, as in this case, failure of preferential independence indicates that there is a higher order value (in this case, whether there are facilities for exercise), and if the two preferentially dependent criteria are replaced with the single more fundamental one, the problem is resolved. For a discussion of models which make implausible preference independence assumptions in the health domain, see Morton (forthcoming).

### **Step 6. Score the options on the criteria**

MAVT involves making numerical assessments of value and of relative importance. Sometime this can be hard for people to do because they are used to thinking of numbers as representing data about things which are “out there in the world”. This is the wrong way to think about the numbers which are used in MAVT: numbers are used but as part of a language to express how people feel about their values. Questions which are mathematically equivalent from the point of view of the multicriteria model can often be experienced psychologically as being quite different (Morton and Fasolo, 2009). For this reason it is often useful to have different ways to ask MAVT elicitation questions: I will review some of these different ways in this subsection.

It is conventional to use a scale bounded by 0 and 100 within each criterion to score options. The performance levels which are defined as 0 and 100 are called the lower and upper reference points. In single choice problems, a common approach is to set the worst performance level in each criterion as 0 and the best as 100; an alternative approach is to anchor the scale at 0 by some absolute idea of a “neutral” level of performance and at 100 by some absolute idea of a “good” level of performance. In problems other than single choice problems, it is good practice to set the *do nothing* baseline level of performance equal to zero (this may mean

that some options have negative scores). This is required in order to ensure that the value of two options together (against the baseline) is equal to the sum of the individual values of the options (against the baseline) (see Morton, 2015, for more details).

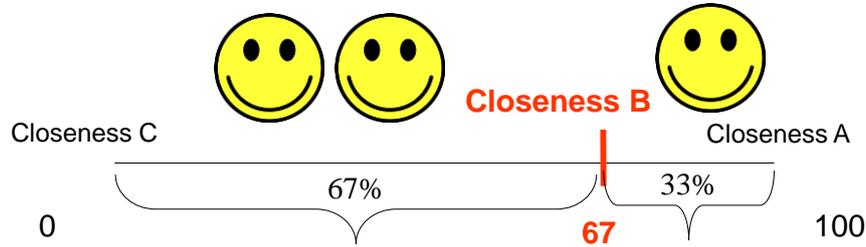
Once 0 and 100 have been assigned, it remains to score the remaining options. The scores should have a preference intensity interpretation. This means, they should represent how intensely option  $a$  is preferred to be  $b$  relative to how intensely  $c$  is preferred to  $d$ . For instance, if the difference between the scores of  $a$  and  $b$  is 40 points, and the difference between the scores of  $c$  and  $d$  is 20 points, then  $a$  is preferred to  $b$  twice as strongly as  $c$  is preferred to  $d$ .

To actually establish the value scores of these intermediate points, it is helpful to have multiple ways to help the elicitee access their values. For example, one can ask the elicitation question as follows:

Suppose you living in a house in Zone C and you woke up one morning to find your house had been moved to Zone A. You would feel happy, right? Fix in your mind how happy you would feel... Now, suppose instead of your house moving from Zone C to Zone A, it only moves to Zone B.... You would still feel happy, but you would feel less happy, right? Now, can you tell me how big is the second amount of happiness as a fraction of the first amount of happiness?

If the answer to this line of questioning is, say, “I would feel two thirds as happy”, then the value Zone B should be 67 (on a scale where Zone C is zero and Zone A is one hundred). It is normally to do “consistency checks” on such number. For example, if Zone B does indeed have a score of 67, this means that a move from Zone C to Zone B should give twice as much happiness as a move from Zone B to Zone A. It is generally worth checking out with the elicitee whether this does indeed correspond to how they feel about the options.

Often there is a certain amount of initial resistance to expressing such quantitative judgements. The elicitor should give the elicitee time to surface the qualitative arguments which may support a judgement of preference intensity. To facilitate the expression of a preference judgement, it is often useful to draw measurement scales and or different numbers or smileys to represent different degrees of happiness (see Figure 5). In group settings, a useful way to get a discussion going is to ask each member of the group to privately assess a score and then compare and discuss differences.



**Fig. 5. Assigning a score for the intermediate level of closeness**

One way to avoid the reluctance which many people feel to putting numbers on their feelings is to ask not for quantitative scores but for qualitative statements about strength of preference. This is the approach of Analytic Hierarchy Process or AHP approach and the MACBETH approach (see Belton and Stewart, 2002 for a presentation of both approaches in a comparative context). MACBETH is fully compatible with the MAVT paradigm, whereas AHP has been criticized in the decision analysis literature on the grounds that it can lead to rank reversals (Dyer, 1990).

A screenshot from the MACBETH software is shown in Figure 6. In the software, options are arranged in a matrix, and elicitees are invited to make statements about the qualitative strength of preference between a number of different pairs. For example, the elicitee may state that the difference in preference in terms of cost between House 4 (the cheapest) and House 1 (the most expensive) is “extreme”, whereas the difference between House 4 and House 3 (the next cheapest) is merely “very weak”. The MACBETH software will then construct a value scale placing the options at appropriate points on the scale, by using linear programming optimisation in which the variables are the scores. The software also facilitates other forms of analysis. In particular the software has an inbuilt function which performs consistency checks on the matrix of comparisons (to identify situations where e.g.  $a$  is strongly preferred to  $b$  and  $b$  is strongly preferred to  $c$  but  $a$  is only weakly preferred to  $c$ ) and suggests how consistencies can be resolved. For further introduction to MACBETH, see Bana e Costa and Chagas, 2004; and Bana e Costa et al, 2011).

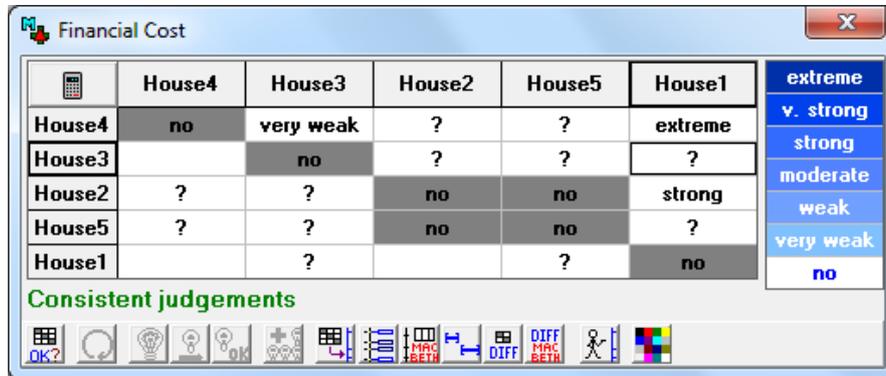


Fig. 6. Establishing value scores for cost: the MACBETH approach

Table 3 shows some possible scores in the house choice problem, with the lower reference point set as the worst level of performance and the upper reference point set as the best level of performance.

	Criteria			
House	Financial Cost	Closeness	Character	Size
1	0	100	100	67
2	50	70	100	67
3	95	0	0	100
4	100	0	0	0
5	50	70	0	67

Table 3. Attribute scores for the house choice problem

Note that the criterion-specific scores as depicted in Table 3 are simply vectors of numbers. If the underlying attribute is continuous (e.g. money, quantity of emissions etc), it may be possible to draw a value function. A value function captures graphically how incremental value changes as the level of performance changes. Figure 7 shows a possible value function for cost. Note that this value function is

decreasing, capturing the idea that lower costs are preferred to higher ones; it is also non-linear, capturing the idea that the decision maker cares more about an increment of £40,000 in cost from a base of £180,000 than from a base of £120,000 (i.e. the difference in value between £120,000 and £160,000 is about 30 whereas the difference in value between £180,000 and £220,000 is 50).

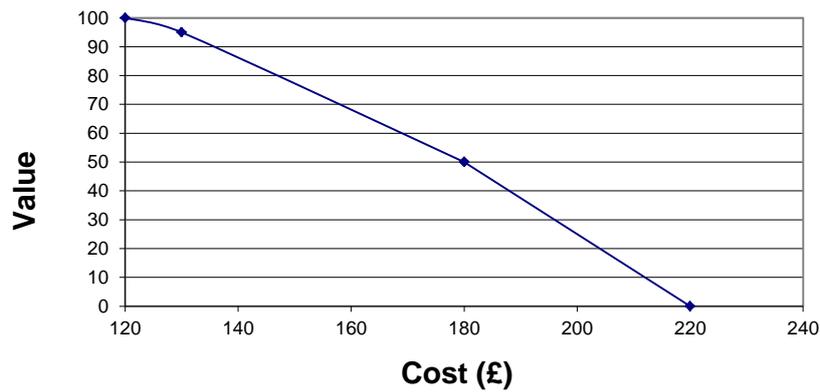


Fig. 7. A possible value function for cost

One natural way to elicit a value function is to use the *bisection method*. When using this method, one asks the elicitee to find a price point  $x$  such that a reduction in cost from the highest level (£220K) to  $x$  yields the same amount of value as a reduction in cost from  $x$  to the lowest price level (£120K). Since the most preferred price point has a score of 100 and the least preferred a score of 0,  $x$  should therefore have a score of 50. By iterating this procedure, price points corresponding to value scores of 25, and 75 can be found, and then corresponding to 12.5, 27.5, 62.5 and 87.5.... to any required degree of articulation.

It should be noted that value functions are quite different from performing an (often arbitrary) normalization of the attribute scales. Normalizations are an automatic mathematical operation that does not represent preferences. A value function represents preferences and therefore must result from an elicitation process.

### Step 7. Weight the criteria

Once scores have been established, the next step is to weight the criteria. The rea-

son for weighting is that although options have been scored on individual criteria, criteria scales are not commensurable: a unit of value on one criterion scale is not the same as a unit of value on another scale. It is as if the options had been valued in terms of different currencies: UK pounds, euros, US dollars, etc.

Weighting thus sets the “exchange rates” between the different criteria. It is critical to do weighting properly as this is what distinguishes MAVT from *ad hoc* approaches. In *ad hoc* approaches, people often set weights by asking questions such as “how important is this criterion relative to that criterion?”. Although people can answer such questions, the questions themselves are meaningless (Morton and Fasolo, 2009). In MAVT, the weighting questions are phrased in terms of increments on different scales.

To see this the difference, consider the question “Which is more important, saving money or saving lives?”. This question as posed is ill-formed. However, the question of how much one is prepared to pay to correct implement a safety feature which will save on average such-and-such a number of lives is a well-formed question. MAVT relies on questions of this latter type.

The most popular method of weighting in MAVT depends on the concept of *swings*. A swing is typically defined as an increase in performance from the level of performance associated with the lower reference point on some criterion to the level of performance associated with the upper reference point. A weight reflects the value of a swing, i.e. the value of improving an option which performs at the lower reference point level on some criterion, so that it performs at the upper reference point level on that criterion. Conventionally the weight of the most valued swing is set as 1 and the weights of the other swings are set as fractions of the most valued swing.

Just as in scoring, swing weighting involves asking questions about hypothetical changes in options. The following question can be used to produce a ranking of the swings.

Imagine you are going to buy a house which has the worst performance levels on all criteria (it costs £220K), is situated in Zone C, has no character, and is only 500 square feet in size. One day, your fairy godmother appears and offers to grant you some wishes. She is unsure how many wishes she has to grant and asks you to prioritise. You may reduce the cost to £120K, change the location from Zone C to Zone A, bestow the flat with character, and increase the size to 700 sq feet. Which do you choose first, which second, which third, and which fourth?

This procedure generates a ranking of the swings. (In our case, suppose the ranking is Financial Cost, Closeness, Size and Character.) The next step is to ask the

“how much” question: how much do you like the second swing as a proportion of how much you like the first? how much do you like the third swing as a proportion of how much you like the first? how much do you like the fourth swing as a proportion of how much you like the first?

The principles behind asking and answering such questions are exactly the same as, and build on the scoring questions: allow elicitees time to reflect and debate, visualize, and make consistency checks to ensure that results “feel right”. The MACBETH software can also be used for weighting, by eliciting qualitative statements about strength of preference (“extreme”, “very strong”, etc.) between the possible swings. a particular advantage of this software is that it also incorporates dominance checks which can supplement quantitative scores by showing how strong the evidence is that one option is overall more highly ranked than another.

Table 4. shows swings and associated swing weights for the house choice problem.

	Criteria			
	Financial Cost	Closeness	Character	Size
Worst performance level	220	3	No	500
Best performance level	120	1	Yes	700
Swing	220→	3→	no→	500→
	120	1	yes	700
Unnormalised Swing weight:	1.00	0.85	0.30	0.50

Table 4. Swings and weights for the house choice problem

As in the case of scoring, where attributes are continuous, this allows an alternative procedure for weighting, called tradeoff weighting. The idea in tradeoff weighting is to adjust the more preferred swing until it yields as much value as the less preferred swing. The concept is depicted in Figure 8. Suppose we have two options, Option 1 which is cheap but poky (£120K, 500 square feet) and Option 2 which is roomy but expensive (£220K, 700 square feet). We like both of these flats better than an expensive and poky flat (£220K, 500 square feet, called the

“nadir”). Moreover, we know from the answer to our fairy godmother question that we would prefer the Option 1 to Option 2: Financial Cost is our most valued swing.

Now we want to ask the “how much” question. But instead asking it directly, we can ask in the following way. Suppose that I adjust Option 1 downwards, in the direction of the nadir, by increasing the price. At some point, Option 1 will cease to be better than Option 2, and become first indifferent and then worse. By locating the point at which indifference occurs, I can find a weight for Size in terms of Financial cost. The reasoning works as follows: I read the price level of the indifference price (£180K, say), and look it up on my value function for cost. From this I see that a price of £180K as compared to £120K is worth 50 value points, measured on the scale of the value function for cost. Since the value of the swing from 500 to 700 square feet is 100, measured in the scale of the value function for size, if I want to express the value of square footage in a way which is commensurable with the value of cost, I must divide the value scores for size by 2, ie use a weight of 0.5.

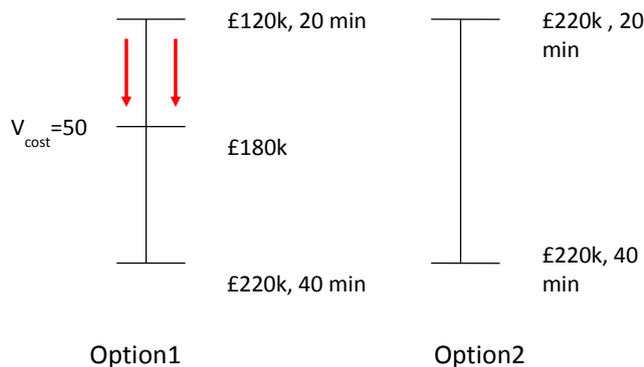


Fig. 8. Sketch of the procedure for tradeoff weighting

### ***Analyzing the model***

#### **Step 8. Compute overall rankings**

Given the scores and weights, and if the options and criteria have the properties outlined in Steps 4 and 5, then it is legitimate to compute an overall value score

for each option  $a$  using the following formula

$$v(a) = \sum_j w_j v_j(a)$$

where  $w_j$  is the weight of criterion  $j$  and  $v_j(a)$  is the score of option  $a$  on criterion  $j$ . This provides a ranking of all options, and can be used to identify the best option, or  $k$  best options.

It should be noted that “weight of criterion  $j$ ” is something we often say in common language, but more formally it should be called “the scaling constant associated with value function  $v_j$ ”. Since these weights might not match the decision makers’ intuition (e.g., “how come safety has such a low weight?”) it might be useful to communicate it as the weight of value function  $v_j$  (or the weight of swing  $j$ ). Bana e Costa et al (2008) present an interesting and instructive application where particular attention was paid to designing the swing weighting procedure so that the swing weights corresponded closely to the decision makers’ natural prior understanding of criterion importance.

Sometimes where there multiple options can be implemented together and there is a concern for value for money, an alternative formula

$$vfm(a) = \frac{\sum_j w_j v_j(a)}{c(a)}$$

may be used, where  $j$  indexes the criteria on the benefit side of the value tree only, and  $c(a)$  is the cost of option  $a$  (excepting the “do-nothing” option which has cost of zero). This formula has the advantage that ordering the options according to this formula and proceeding down the list until the budget is exhausted, will give a good solution to the budget allocation problem, especially if there are many options. For more ideas on how to deal with this particular *problématique*, see Salo et al (2011) and Morton et al (2016).

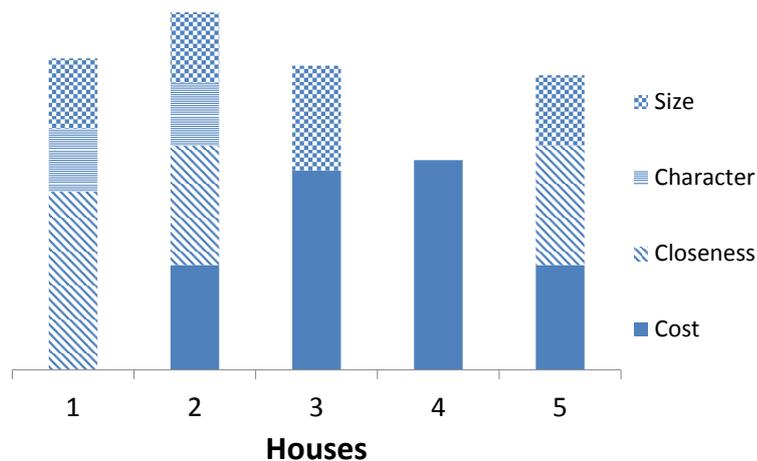
### **Step 9. Conduct sensitivity analysis**

Often people consider that a multicriteria analysis is complete when they have scored options and weighted criteria and arrived at a ranking of options. Nothing could be further from the truth. The aim of MAVT is not to find the “right answer” – where there are conflicting objectives, no right answer exists – but to ena-

ble decision makers and stakeholders to explore the problem and come to a considered decision. Sensitivity analysis involves varying scores or weights in an interval and noting the impact on the model results. Sensitivity analysis can reveal how important uncertainties or disagreements (such as those identified in Steps 6-7) are on the final results.

I now present three sensitivity analysis displays for the house choice problem: the stacked bar chart, the Pareto chart, and the parameter-wise sensitivity analysis.

Figure 9, the stacked bar chart, shows the composition of aggregate value for the different options. From this it can readily be seen what options are cheap (a lot of the value of Houses 3 and 4 is due to their strong performance on the cost criterion). House 4 in particular has nothing to recommend it except that it is cheap. House 1 gets a great deal of value from closeness and if the elicitee really cared about closeness she would choose this option, but the winner seems to be – given these scores and weights – House 2 which has the advantage that it is a good all-rounder, with cost, closeness, character and size reasons to recommend it.



**Fig. 9. Stacked bar chart for the house choice problem**

Figure 10 shows a Pareto chart. In this display, the scores for Financial cost are plotted against a weighted combination of the scores of all other criteria (“Benefits”). Houses on the frontier of the green area are efficient in the sense that for each house, that there is some assignment of weights to “Costs” and “Benefits” which makes that house the highest valued house. House 1 is the point on the vertical axis (it has all the benefits but is expensive); House 4 is the house on the

horizontal axis (it has no benefits but is cheap) and Houses 2 and 3 are the points on the curve, both representing a compromise between costs and benefits. Note that House 5 is not efficient in this display. This is a consequence of house 5 being dominated. It is however possible for an option to be not efficient even if it is not dominated.

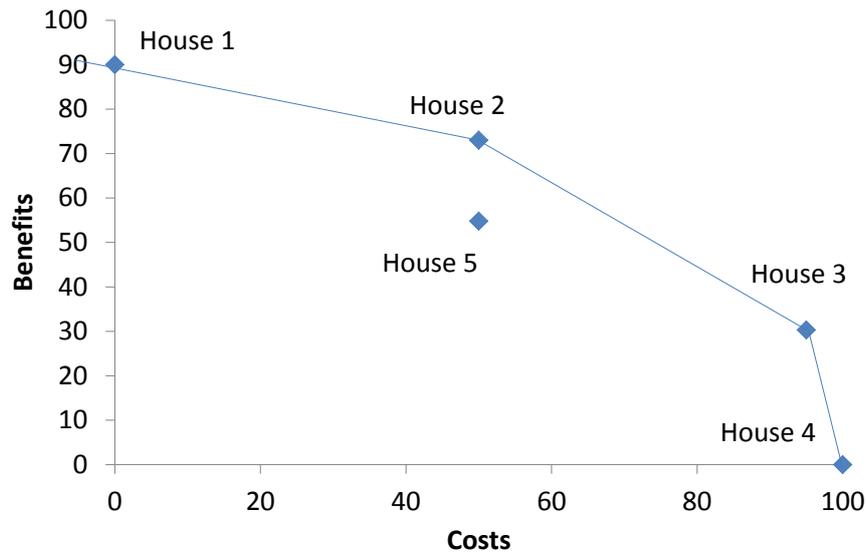
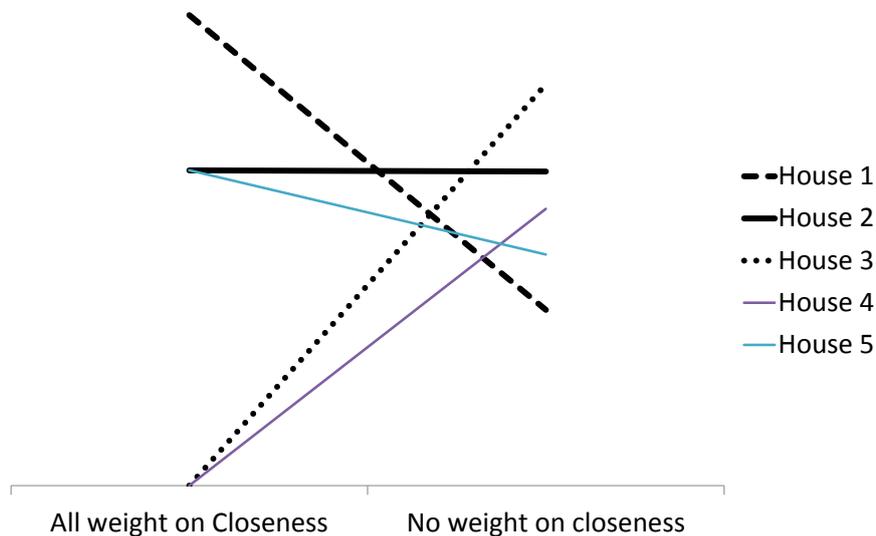


Fig. 10. Efficient front display for house choice problem

Figure 11 shows a parameter-wise sensitivity analysis for the criterion Closeness. This display shows how the valuation of the options changes as the weight on Closeness is varied relative to the weight on the other criteria whilst holding the relative weights on the Benefit criteria fixed. From this display it can be easily seen that: House 1 is a good option if Closeness is high weighted relative to the other criteria; House 2 is a good option if Closeness is intermediate weighted relative to the other criteria; and House 3 is a good option if Closeness is low weighted relative to the other criteria. The other two options do not, for this analysis and given these numbers, make it into the running.



**Fig. 11. Parameter-wise sensitivity analysis: value of options varying weight on Closeness for house choice problem**

Although sensitivity analysis in MAVT can be done using spreadsheets, it is often more efficient to use software (for example, Hiview or VISA or MACBETH or WISED for single choice problems; Equity or PROBE for multiple choice and budget allocation problems) as these softwares have built-in sensitivity analysis tools. The technical literature has a wider selection of ideas and tools for performing sensitivity analysis (e.g. Rios Insua and French, 1991; Dias and Clímaco, 2000; Borgonovo and Plischke, 2016) but these have not yet generally been incorporated in professional commercially-available user-friendly software.

### ***Troubleshooting***

In this section of the chapter, I consider some commonly occurring problems in applying MAVT in practice, and suggest some ways to approach such problems.

1. ***There are too many options.*** Consider using a small number of screening criteria to establish a shortlist (e.g. would this option require new legislation to implement? Would it cost more than £x?). If several options are similar (e.g., small variations), consider evaluating only one from each group/cluster and, if it turns out to be among the best, only then evaluate the ones similar to

it.

2. ***There are not enough options.*** Look for solutions which other organisations have implemented when faced this or similar decisions. Consider holding a brain storming session. Consider enlarging the scope of the analysis. E.g., someone suggests you do voluntary work at an hospital two hours per week. Instead of considering the alternatives “yes” and “no”, you might consider the alternatives are how many hours will you devote to this organization, or consider the alternatives are different organizations where you could do voluntary work, adding the “none” option and perhaps other ways to use your free time.
3. ***The options do not seem to be comparable.*** Come up with a description of what options should be (e.g. “facilities”; “development plans”). Restructure the options by merging some or deleting some.
4. ***The options cannot be evaluated independently of each other.*** Consider restructuring the options (e.g. merging options which have a dependence relation; assuming that one option on which several others depend will be done). Alternatively, consider using more complicated analytic techniques such as mathematical programming. Consider using Portfolio Decision Analysis methods, see Salo et al (2011) and Morton et al (2016).
5. ***There are too many criteria.*** Look for criteria which are redundant, ie which duplicate each other; which do not discriminate between options. Consider merging similar criteria into higher level criteria.
6. ***There are not enough criteria.*** Look for criteria which other organisations have implemented when facing this or similar decisions. Consult published documents such as strategic plans. Consider holding a brain storming session. Consider what important attributes might differentiate two alternatives that are similar on the criteria you already have.
7. ***The criteria are not preferentially independent.*** Consider restructuring the criteria (e.g. merging two criteria which are dependent because they are alternative ways of achieving some higher order goal). Use a non-additive value model.
8. ***Participants don’t understand scoring and/ or weighting.*** Use software, or draw pictures on flip charts to help participants visualize. Ask questions in different ways, using the different questioning modes listed in this chapter. Use analogies to communicate weight and scale concepts (e.g. exchange rates; metric and imperial scales; Celsius and Fahrenheit). Build models in real-time allowing to observe how outputs change as inputs also change.
9. ***The overall values don’t “feel right”.*** Ask yourself and your decision maker why the answers don’t feel right. Is there a missing criterion? Do you really believe the scores and weights? Use sensitivity analysis to explore the model.

10. ***There isn't enough time to do everything properly.*** One option is to proceed with incomplete information and check what is robust to save time, see Dias (2007). However, a decision analysis can take various forms – from a quick back-of-the-envelope analysis in an hour or two to workshops spread over several days. Use the time you have, and be realistic about what you can achieve.
11. ***The decision makers or significant parties do not have time to participate.*** Do not demand very exact answers (e.g. are you sure the score is 50 and not 49 or 51?). Often, sensitivity analysis shows that small imprecisions do not matter (“flat maxima principle” of von Winterfeldt –and Edwards, 1986). Remind the decision maker that the analysis is a tool to help them structure and think through the decision, not something which will or should try to take the decision for them. If time is an issue, not everyone has to be involved in every stage of the decision process (for example a small working group may define criteria and options which can be scored by a larger group).
12. ***The decision makers or significant parties are afraid of losing control of the decision.*** Not everyone has to be involved in every stage of the decision process (for example weights may be defined by the management team or by a single client). Control is not absolute in any case, and often decisions which are arrived at by a non-transparent process are hard to implement because of stakeholder resistance.
13. ***The decision makers do not agree on some inputs.*** Build different models in parallel or use incomplete information they agree with (e.g., they do not agree on the weights  $w_1$  and  $w_2$ , but agree that  $w_1 > w_2$ ). Assess what common results can be obtained. Often, different inputs lead to the same outputs.
14. ***The decision makers refuse the idea of trade-offs*** (e.g., harm to the environment vs. harm to health vs. costs). This may be caused by options with unacceptable performance on key criteria that the decision makers feel cannot be compensated by good performance in another criterion. In such cases, consider removing these unacceptable alternatives. Otherwise, using MAVT might not be the best option and outranking methods (Chapter 14) or other approaches might be appropriate to such type of decision makers.

## Concluding Remarks

The founding texts of MAVT (Keeney and Raiffa, 1976; von Winterfeldt and Edwards, 1986) are now 40 and 30 years old respectively. Although younger by sev-

eral decades (or centuries, depending on how one counts) than probability theory, MAVT can therefore also be considered to be a mature technology.

Is it a successful technology? Considered in its broadest sense, the answer has to be yes: the scoring and weighting approach is (as far as one can tell) very widely used in applied settings, such as R&D prioritization and procurement. However, many users of scoring and weighting have never heard of MAVT, and are unaware that that a body of theory-based knowledge exists about how to perform elicited scores and weights. To some extent this is also true of probabilistic modelling also. However, much of the use of probabilistic concepts is mediated by software such as spreadsheet simulation packages and such software provides an easy bridge for users to learn more about probabilistic concepts. Software based around MAVT concepts has not (yet) enjoyed such widespread success.

Like the authors of the Chapter 9 (in their Discussion section) I see huge potential for MAVT methods in an increasingly digital and data-rich world. Currently if one is shopping online for hotel rooms or flights, the search engines allow one to rank order options on the basis of holistic assessments, or on the basis of individual criteria, but provide little in the way of support for locating the option which has the ideal balance of attributes given one's preferences. It is plausible that increasingly demanding online consumers will at some point start to demand better decision support to enable them to cope with the vast mass of undigestible information which is regularly served up to them.

However, the original promise of MAVT as a rigorous yet transparent framework for choice was to help support big policy decisions as well as small personal ones. There are some signs in some domains that multicriteria methods are meeting with increasing favour. In the area of health technology regulation and assessment, for example, there has been a recent upsurge in interest in the use of multicriteria methods to support medicines regulation and reimbursement decisions (Thokala et al, 2016; Marsh et al, 2016). However, there is still a substantial gap between the potential for the formal use of MAVT to beneficially support substantial decisions in government and business, and actual current practice. Hopefully that gap will close in the years and decades ahead.

**Acknowledgements.** I would particularly like to acknowledge the many helpful comments of Luis Dias, many of which have been directly incorporated in the text.

## References

- Ackermann F, Eden C (2011) Strategic Management of Stakeholders: Theory and Practice. *Long Range Plann* 44 (3):179-196
- Bana e Costa CA, Chagas MP (2004) A career choice problem: an example of how to use macbeth to build a quantitative value model based on qualitative value judgments. *Eur J Oper Res* 152 (2):323-331
- Bana e Costa CA, de Corte J-M, Vansnick JC (2012) MACBETH. *International Journal of Information Technology & Decision Making* 11 (2):359-387
- Bana e Costa CA, Lourenço JC, Chagas MP, Bana e Costa JC (2008) Development of Reusable Bid Evaluation Models for the Portuguese Electric Transmission Company. *Decis Anal* 5:22-42
- Belton V, Stewart TJ (2002) *Multiple Criteria Decision Analysis: an integrated approach*. Kluwer, Boston
- Bond, SD, Carlson, KA, Keeney, RL(2010) Improving the Generation of Decision Objectives. *Decis Anal* 7: 238–255. doi:10.1287/deca.1100.0172
- Borgonovo E, Plischke E (2016) Sensitivity analysis: A review of recent advances. *Eur J Oper Res* 248 (3):869-887
- Bryson JM (2004) What To Do When Stakeholders Matter Stakeholder Identification and Analysis Techniques. *Public Management Review* 6 (1):21-53
- Dias LC (2007) A note on the role of robustness analysis in decision-aiding processes. In: B. Roy, M. Ali Aloulou, R. Kalai (eds), *Robustness in OR-DA*, Annales du LAMSADE, No. 7, Université-Paris Dauphine, pp. 53-70
- Dias LC, Clímaco JN (2000) Additive aggregation with variable interdependent parameters: the VIP analysis software. *J Oper Res Soc* 51 (9):1070-1082
- Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW (2015) *Methods for the Economic Evaluation of Health Care Programmes*. Oxford University Press, Oxford
- Dyer JS, Sarin RK (1979) Measurable multiattribute value functions. *Oper Res* 27 (4):810-822
- Dyer JS (1990) Remarks on the Analytic Hierarchy Process. *Manage Sci* 36 (3):249-258
- European Food Safety Authority (2014) *Guidance on Expert Knowledge Elicitation in Food and Feed Safety Risk Assessment*. *EFSA Journal* 12 (6)
- French S (1986) *Decision Theory: an introduction to the mathematics of rationality*. Ellis Horwood, Chichester
- French S, Rios Insua D (2000) *Statistical Decision Theory*, vol London. Kendall's Library of Statistics. Arnold,
- Goodwin P, Wright G (2014) *Decision analysis for management judgement*. 5th edition. Wiley, Chichester
- Howard RA, Abbas AE (2016) *Foundations of Decision Analysis*. Pearson, Harlow, Essex
- Keeney RL (2002) Common mistakes in making value trade-offs. *Oper Res* 50 (6):935-945
- Keeney RL, Raiffa H (1976) *Decisions with multiple objectives: preferences and value tradeoffs*. Wiley, Chichester
- Köbberling V (2006) Strength of preference and cardinal utility. *Economic Theory* 27 (2):375-391
- Krantz DH, Luce RD, Suppes P, Tversky A (1971) *Foundations of Measurement Vol 1*. Academic Press, New York
- Marsh K, M II, Thokala P, Baltussen R, Boysen M, Kalo Z, Lonngren T, Mussen F, Peacock S, Watkins J, Devlin N (2016) *Multiple Criteria Decision Analysis for Health Care Decision Making- Emerging Good Practices: Report 2 of the ISPOR MCDA Emerging Good Practices Task Force*. *Value Health* 19 (2):125-137. doi:10.1016/j.jval.2015.12.016
- Matheson JE, Howard RA (1983) An introduction to decision analysis. In: Howard RA, Matheson JE (eds) *The Principles and Applications of Decision Analysis*. SDG, Menlo Park,
- Miyamoto JM, Wakker PP, Bleichrodt H, Peters HJM (1998) The zero-condition: A simplifying as-

- sumption in QALY measurement and multiattribute utility. *Manage Sci* 44 (6):839-849
- Franco LA, Montibeller G (2010) Facilitated modelling in operational research. *Eur J Oper Res* 205 (3):489-500
- Morton A (2015) Measurement issues in the evaluation of projects in a project portfolio. *Eur J Oper Res* 245 (3):789-796
- Morton, A (forthcoming). Treacle and Smallpox: Two Tests for Multicriteria Decision Analysis Models in Health Technology Assessment. *Value in Health*.
- Morton A, Fasolo B (2009) Behavioural Decision Theory for Multi-Criteria Decision Analysis: a guided tour. *J Oper Res Soc* 60 (2):268-275
- Morton A, Keisler J, Salo A (2016) Multicriteria Portfolio Decision Analysis for Project Selection. In: Ehrgott M, Figueira JR, SGreco (eds) *Multiple Criteria Decision Analysis: State of the Art Surveys 2nd Edition*. Springer, New York,
- Phillips LD (2007) Decision Conferencing. In: Edwards W, Miles RF, Von Winterfeldt D (eds) *Advances in decision analysis: from foundations to applications*. CUP, Cambridge,
- Pliskin JS, Shepard DS, Weinstein MC (1980) Utility functions for life years and health status. *Oper Res* 28 (1):206-224
- Rios Insua D, French S (1991) A framework for sensitivity analysis in discrete multi-objective decision-making. *Eur J Oper Res* 54:176-190
- Roy B (1985) *Méthodologie multicritère d'aide à la décision*. Economica, Paris
- Salo A, Keisler J, Morton A (eds) (2011) *Portfolio Decision Analysis: methods for improved resource allocation*. Springer, New York
- Thokala P, Devlin N, Marsh K, Baltussen R, Boysen M, Kalo Z, Longrenn T, Mussen F, Peacock S, Watkins J, Ijzerman M (2016) Multiple Criteria Decision Analysis for Health Care Decision Making-An Introduction: Report 1 of the ISPOR MCDA Emerging Good Practices Task Force. *Value Health* 19 (1):1-13. doi:10.1016/j.jval.2015.12.003
- von Winterfeldt D, Edwards W (1986) *Decision analysis and behavioral research*. CUP, Cambridge