Monograph:
Pearman, A.D. (1989) Developing a Multicriteria Model for Use as a Highway Assessment Technique. Working Paper. Institute of Transport Studies, University of Leeds , Leeds, UK.

Working Paper 281

Institute of Transport Studies
University of Leeds

Published paper

Pearman, A.D. (1989) *Developing a Multicriteria Model for Use as a Highway Assessment Technique.* Institute of Transport Studies, University of Leeds. Working Paper 281

Working Paper 281

September 1989

DEVELOPING A MULTICRITERIA MODEL FOR USE AS A HIGHWAY ASSESSMENT TECHNIQUE

A D Pearman

School of Business and Economic Studies

This work was sponsored by the Economic and Social Research Council

DEVELOPING A MULTICRITERIA MODEL FOR USE AS A HIGHWAY INVESTMENT PRIORITY ASSESSMENT TECHNIQUE

1. INTRODUCTION

This paper is concerned with the development of a simple multicriteria model for use as a priority assessment technique (PAT) by local authority transport planners faced with the problem of identifying which of a range of highway investment proposals should be implemented. The project of which it forms a part has involved three main phases:

Phase I a review and critique of PATs developed by British local authorities;

Phase II the application of a representative sample of PATs to a set of six highway schemes, together with an analysis of the different scheme rankings which emerged;

Phase III the construction, based upon the experience of phases I and II together with knowledge of recent developments in multicriteria analysis, of a computer-based PAT.

An account of the outcome of the first two phases of the project is given in Simon et al (1987); more detailed information is available in Simon (1986a,b; 1987).

2. BACKGROUND

The use by local planners of formal, quantitative priority assessment techniques to help rank and select for implementation competing low cost transport projects is common, in the UK, USA and elsewhere. The typical institutional framework within which such PATs operate has:

- a large number of candidate schemes, with aggregate costs considerably in excess of budget limits;
- schemes varying substantially in cost, in design and in the nature of their likely impacts;
- limited time and manpower resources available for assessment and evaluation;

- a decision process which is often subject to political influence or the pressures of public opinion;

- impacts which it is often not practicable to assess by objective measurement;

- the need to repeat the selection process, typically on an annual basis, with new projects added to the pool and possibly with a new set of decision makers.

Since the early 1970's, a substantial number of UK local authorities have developed, more or less independently, assessment frameworks to help with the task of creating a logical and defensible annual investment programme (see Simon, 1986a for details). Most of the techniques concerned are "points-scoring methods", in which each candidate scheme is assessed against a series of attributes and schemes are prioritised according to their aggregate points score (sometimes weighted to reflect the relative importance attached to different attributes). The outputs from the application of a PAT by no means finalise the investment programme. Their primary use is within planning offices, as one input to the prioritisation process, although they can also be used as a basis for discussion between officers and elected members about the choice of schemes for implementation. There is substantial variation in the level of detail at which different authorities' PATs work, from a minimum of four attributes to a maximum of 43 in the set of techniques we examined. PATs may be applied very early in the planning process, as a screening device, or later, at a stage much closer to the development of a final programme.

The picture in the USA is broadly similar. Different local highway authorities have experimented over the last decade or so with a wide range of techniques. Since the US institutional framework for highway planning is different from that in the UK, not all the problems, nor all the techniques, are directly relevant to PAT development in the UK. Useful references include Transportation Research Record numbers 1116 and 1124, Harness and Sinha (1983) and TRB (1984). For information about current practice in continental Europe Leleur (1985) and Himanen (1987) are helpful.

Our assessment of the range of PATs being used by local authorities in Great Britain led us to conclude that it would be a worthwhile exercise ourselves to construct a PAT to provide a logical, consistent and comprehensive framework which local authorities could employ. From the outset, we restricted our attention to highway projects. We also concentrated in the first instance on providing a means of evaluating the predicted effects of candidate schemes, rather than assessing the relative importance of identified highway "problems". The latter is an important question related to scheme assessment, since many authorities not unnaturally wish to identify and do something to ameliorate the most extreme perceived problems in their highway network. Issues raised by the possibility of undertaking both problem severity assessment and scheme evaluation within a single ranking exercise are discussed in Section 5.

In designing a PAT, our aim was to construct a computer-based model, building on existing best practice among the local authorities, but also recognising that the prioritisation exercise was a form of multiple criteria decision making (MCDM). Any procedure we devised should exploit the substantial growth in understanding in recent years of how multicriteria evaluation and choice can be aided by formal quantitative techniques.

From the point of view of transport planning practice, we identified the following as capabilities that our PAT should have:

(a) to store and present information about projects in a straightforward fashion, e.g., in a matrix/framework with rows corresponding to different attributes and columns corresponding to different projects;

(b) to set out a comprehensive list of possible impacts which ought to be taken into account in assessing local highway improvement projects, together with suggested attribute scales for measuring the impacts;

(c) to permit nonetheless a degree of flexibility as to what impacts are assessed and how they are measured in order to allow different local authorities to tailor the PAT to their own needs;

(d) to provide a default set of weights for the attributes, but also to enable the user to create his/her own set of weights;

(e) to present information about the different projects graphically and in other readily digested formats;

(f) to facilitate sensitivity testing on both attribute weights and project scores, in the latter case recognising that limited time and manpower availability may restrict the extent and accuracy of assessment of individual schemes;

(g) to permit projects for which only preliminary information is available to be assessed in the same framework alongside projects which are more fully specified;

(h) to be consistent with the possibility of assessing problem severity within broadly the same structure as scheme performance.

In seeking to construct a PAT that met with these requirements, we restricted our attention to highway schemes in the broad cost range £25K to roughly £2M. Smaller schemes would be unlikely to receive much in the way of formal appraisal; substantially larger ones would in all probability be the subject of a wider public enquiry process, with different decision procedures. Given the importance of cost–benefit analysis to transport project evaluation in general and hence its familiarity to transport planners, we attempted to ensure that the presentation of results was broadly consistent in style with the output of a CBA, even though the importance of environmental and planning considerations in local scheme prioritisation effectively prohibits the use CBA for the priority assessment process itself.

3. MULTIPLE CRITERIA DECISION MAKING AND HIGHWAY PROJECT PRIORITY ASSESSMENT TECHNIQUES

The intention of this section is to give a brief account of the main techniques of multiple criteria decision making, followed by an assessment of what the general MCDM literature implies for the construction of a PAT based on MCDM principles.

In 1772 Benjamin Franklin wrote to Joseph Priestley (Wilcox, 1975):

Dear Sir, London Sept. 19. 1772

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

B FRANKLIN

Dr Priestly

Since that time and especially in the last couple of decades there has been an enormous growth in the theoretical literature relating to formal, more-or-less quantitative techniques for guiding decisions where there are multiple dimensions of impact to take into account. It is clearly important that any PAT oriented towards the type of highway investment schemes of interest to us should take into account what advice the MCDM literature can give. At the same time, it is perhaps worth noting that some of the PATs that local authorities were found to be using in practice were not so very much more sophisticated than Franklin's technique. While this might say something about the education of transport planners, it might also convey an important message about relationships between the theory and practice of decision making.

For practical purposes, the problem faced in highway scheme priority assessment is one of ranking pre-specified alternatives. There is no project design element; that is, there is no attempt to specify the characteristics of each proposed scheme through the optimisation of some objective function. It is also reasonable to assume that there are no significant interdependencies between the projects being assessed, or, at least, that any such relationships can be handled on an ad hoc basis by the definition where necessary of appropriate combined project packages. For these reasons, the large area of multicriteria work which is primarily concerned with multi-objective programming/decision-making is not directly relevant to our immediate needs. What is relevant is the set of techniques which are normally termed multiple attribute decision making (MADM) models (Hwang and Yoon, 1981, p. 3). These methods are oriented towards prioritisation of fully prespecified alternatives lying within a finite (and usually small) set of possibilities.

MADM methods may be classified according to the amount of information they assume to be available to the decision maker and the nature of that information (Hwang and Yoon, pp.8/9). Based on our survey of existing PATs (Simon, 1986a) the information which is potentially available to guide the highway scheme prioritisation decision is likely to be $\{X_{ij}\}$, an assessment of the level of performance of proposed scheme i on the j^{th} attribute and $\{w_j\}$, an assessment of the relative importance of the attributes. Direct holistic pairwise comparisons of alternative schemes are not generally available and are in any case unlikely to be a helpful basis for prioritisation in the institutional context within which PATs operate.

An important issue is whether the information $\{X_{ij}\}$, $\{w_j\}$ which will form the basic input to the PAT is ordinal or cardinal. For the purposes of developing our PAT, we have taken the view that it will be cardinal. The reasons for doing so are (a) that the discriminatory power between alternatives of those MADM methods that rely purely on ordinal imputs can be very limited; (b) a good proportion of the impacts which existing PATs consider are naturally measured on cardinal scales; (c) trade-offs between different impacts/attributes seem to be important in practice and need to be addressed with as mathematically powerful tools as possible. Some relaxation of the cardinality assumption may, nevertheless, be possible in certain circumstances. This is discussed in Section 5.

Within the set of MADM techniques that produce a prioritisation of alternatives based on cardinal data input, there are still several possibilities. For our PAT development, we rejected from among these the outranking methods deriving from the work of Roy (1985) and also related approaches such as interpretive structural modelling (Janes, 1987). Although outranking models have features that might make their application to highway project prioritisation interesting, from the point of view of the needs of a PAT, they also have some disadvantages: a relatively complex methodology; a limited axiomatic foundation and aspects of their operation which, over and above the $\{X_{ij}\}$ and $\{w_j\}$ are decision-maker or context-dependent and hence not necessarily reproducible. Similarly, at this stage, we rejected the use of Saaty's analytic hierarchy process (Saaty, 1988) as a means of tackling the complete PAT problem, principally because of doubts about its ability in this context to provide straightforward and reliable estimates of $\{X_{ij}\}$. Its possible use to determine the $\{w_j\}$ will be discussed later. Finally, ideal point methods, such as TOPSIS (Yoon 1980) were also rejected, principally on the grounds of lack of adequate axiomatisation.

What remains from this process of elimination is the set of MADM methods falling within the multiattribute utility theory/multiattribute value theory categorisation (Keeney and Raiffa, 1976; Dyer and Sarin, 1979). These operate with cardinal data inputs; are well axiomatised; are reproducible for given data inputs; are relatively transparent in the way they operate; have been used on a good number of public sector applications and also can be implemented at different levels of sophistication, depending on the context of application. For all the above positive reasons, as well as because of the disadvantages we perceived about alternative methods, if was the use of techniques from within this set that we pursued. Moreover, although there clearly are significant uncertainties associated with the consequences of implementing any highway project, the methods we chose to pursue were those based on value theory (implying a deterministic model) rather than the uncertainty-oriented utility theory models. This decision rests on the observation that, for choices of the type that PATs are intended to guide, formalising the degree of uncertainty associated with different projects would be almost impossible in practice. A more promising way of recognising the inherent uncertainties is to encourage sensitivity analysis on a (deterministic) model.

4. DEVELOPING A MULTIPLE ATTRIBUTE DECISION MAKING MODEL FOR USE IN PRIORITY ASSESSMENT

This section describes, against the background sketched out in section 3, the way in which a multiple attribute decision making model was constructed for use in our PAT. Although the process which led to the final choice of model form and its implementation was much less orderly, it will nonetheless help to describe the reasoning behind our model to set it out in terms of a standard framework. The stages which will be discussed are:

- structure the decision problem

- assess the possible impacts of each alternative

- determine the preferences of the decision makers

- evaluate and compare alternatives.

This particular framework is due to Keeney (1982), but is typical of a number of descriptions of how to apply multiattribute analysis. Most such descriptions emphasise the need for iteration between and within sections of the framework as the analyst gradually converges towards a preferred model specification. Thus the somewhat indirect path alluded to earlier as describing the way by which the final form of the PAT multiattribute model was obtained is, in fact, neither specially undesirable nor unusual.

4.1 Problem Structuring

The first main issue to be addressed here as far as a conventional decision analysis is concerned is the generation of alternatives. The second is the specification of the decision-maker's objectives and measurable attributes by which the level of attainment of each of the objectives can be measured.

One respect in which a PAT must differ in practice from the standard theoretical multicriteria model is that the latter is axiomatised on the assumption of modelling the preferences and hence guiding the action of a single individual. It is his/her preferences and judgements which structure and parameterise the model. Clearly a PAT will reflect the judgements of more than one person. It is implicit in the rest of the model development that what is formally an individual decision aid can adequately represent the views of a decision making group. Significant differences of view can be explored through sensitivity testing or through repeat analyses using different weight sets and/or attribute sets.

A second way in which a multicriteria model applied as a PAT will be untypical and outside the strict theoretical framework of multicriteria analysis arises from the task which a PAT is intended to undertake and the institutional framework within which is operates. In its standard presentation, a decision analysis is tailored specifically to a single choice between alternatives whose specifications are known before the project scoring and preference parameters of the evaluation model are fixed. For a PAT, however, the objective is to set up a model adequate to prioritise a range of schemes whose specification is unknown. Its principal intention is to ensure consistent treatment between projects and perhaps between years. Although the nature of the schemes can be anticipated in broad outline, their detail cannot. Also PATs will not usually be implemented by decision analysis specialists. To a good extent they must act as "production line" systems, capable of handling a range of possibilities with minimal case-specific adjustments. All these institutional factors have a bearing on the types of multicriteria analysis which it is practicable to consider for PAT purposes and their implementation in practice.

Although inevitably such institutional and practical considerations imply that the choice model used is unlikely to be precisely correct for each individual prioritisation exercise, it should be remembered, especially bearing in mind the politicised nature of much decision making about local highway expenditure, that a PAT is acting very much as a decision support system and not as a prescriptive device. Moreover, as mentioned in section 2, the alternatives themselves cannot in practice be fully characterised for the choice process. Uncertainty is ignored in the formal model development, as is any possibility of implementing any dynamically staged decision making for what are typically, but not always, small-scale unitary projects.

The second stage of problem structuring involves specifying objectives and attributes. Here again, the special circumstances of PAT application influence the way in which our model development proceeded.

Initially, attention must be given to determining a list of objectives which, between them, specify all factors that are relevant to choice in the circumstances concerned.

Typically, the objectives are structured into a tree hierarchy (see Figure 1); this aids the subsequent decision process in a number of ways (Brownlow and Watson, 1987, pp. 510–12). The higher level objectives set out the overall aims or ends which concern the decision maker. The lower levels in the hierarchy progressively define the higher level ones, effectively specifying the means through which the ends may be attained.

There is some debate about the most effective way in which to construct the hierarchy, top-down or bottom-up. Top-down tree development involves specifying the broad objectives first and then filling in the detailed specific objectives; bottom-up starts by developing a full list of detailed objectives and structures the tree through successive clustering of related lower-level objectives. Adelman et al. (1986) suggest that either approach can yield equally acceptable results. Buede (1986) argues that top-down structuring is most appropriate for strategic decisions, where only the general aims are known and bottom-up for tactical decisions, where the actual alternatives may already be known.

PATs perhaps fit more easily into the mould of strategic decisions; although the individual projects which they analyse are small-scale, the overall objectives implied by the objectives hierarchy have strategic significance (e.g, the balance of emphasis given to environmental as opposed to directly financial considerations). A top-down analysis may also conform more readily with the way in which politicians' general preferences are articulated. In developing the hierarchy for our PAT, elements of both bottom-up and top-down structuring were present.

Through our earlier discussions with local authority planners, we were aware of the range of specific (lowest level) objectives which tended to be employed. At the same time, we were aware that most local authorities classified the lowest-level objectives under higher-level headings. Some, indeed, only evaluated schemes at an aggregate level (several existing PATs used fewer than 10 objectives). Our decision to structure the set of objectives for our PAT into a tree hierarchy was partly influenced by existing local authority practice, but depended more on a number of analytical and technical advantages that a tree structure affords. First, presentationally, a tree structure helps the user grasp quickly the range of objectives which the PAT employs. Secondly, as will be explained in Section 5, a tree structure can facilitate cost-effective assessment of smaller schemes for which the time and manpower input associated with a detailed assessment could not be justified. Finally, a tree structure can help in the process of checking the set of objectives/attributes which has been developed.

The tree structure hierarchy of objectives associated with our PAT is the one shown in Figure 1. With each lowest-level objective in the hierarchy must be associated a measurable attribute to reflect the extent of attainment of the corresponding objective recorded by any particular scheme. A list of the attributes used is given in Appendix 1. The thinking which underlies the choice and scaling of the attributes will not be discussed here (see Mackie et al., 1988). Nonetheless, it is worth noting in passing that

- choice of attributes was influenced by the need to assess individual schemes which did not make excessive demands on manpower or involve unduly expensive or time-consuming monitoring of sites;

- numerically-scaled subjective assessments are used for 11 of the 32 attributes;

- we would expect that some local authorities would wish to amend the chosen set of objectives/attributes to reflect their own circumstances.

The amount of effort that went into the specification of the attributes was less than would be expected in many applications of multiattribute analysis. For example the implications of choosing direct or proxy attributes, natural or constructed scales (Keeney, 1981) were given only limited consideration. In part, this was because there was already a well-established body of practice concerning the evaluation of transport projects which steered the analysis towards the choice of (say) money value of time savings as the attribute/scale through which to measure this aspect of improvements in the efficiency of the highway system. In part, also, it was felt that the appropriateness of the attribute set and the measurement scales was best improved by reacting to the responses of users of the PAT.

Standard practice (Keeney and Raiffa, 1976, pp. 50-3; Von Winterfeldt and Edwards, 1986, pp. 43/4) suggests that, once developed, the set of lowest-level attributes should now be checked against a set of criteria:

- completeness

- operationalisability

- decompability

- non-redundancy

- parsimony

Completeness concerns the extent to which the specified set of attributes/objectives can reflect the degree to which the overall objective (identifying highway schemes which contribute the most effectively to improving local conditions) is attained. Individual attributes need to be both comprehensive and measurable. By comprehensive is meant that the decision maker, knowing the numerical value of the attribute, should thereby have a clear understanding of the extent to which the associated objective is achieved. Clearly, completeness is an ideal to be strived for, but also to be compromised in the light of the practical circumstances of the study. From our thorough review of previous practice and our own knowledge of the field, we believe that the set of lowest-level objectives/attributes given in Appendix 1 adequately meets the completeness requirement in the context of priority assessment. The attribute set should be adequate to differentiate likely projects to the greatest practical extent. Feedback from users of COMPASS (the computer implementation of the multiattribute model, Mackie et al., 1988) will confirm this judgement, or provide a basis for modification of the attribute set.

Operationalisability requires that the attributes must make sense to the decision makers, be employable as a basis for discussion of alternative schemes and be practicable in the particular circumstances of a PAT. Cost-effective measuring of attributes has already been mentioned as an important consideration. As with completeness, whether the attribute set is operational will finally be clarified by the responses of COMPASS users.

Decomposability requires that decision makers are able to "divide and conquer" the overall assessment problem by considering the individual attributes largely independently of each other before recombining them. Similarities between the multiattribute approach and existing PATs suggest that this should prove practicable. This topic will be considered further in section 4.3.1.

Non-redundancy is the requirement to strive to avoid double-counting. The hierarchical structure of the objectives tree is intended to diminish the danger of redundancy. However, especially in view of the subjective judgement scales that are predominant in some sections of the tree, care needs to be exercised. For example, in assessing highway characteristics (attributes A2.1 - A2.4) it may be difficult to assess independently the four separate characteristics; in assessing planning/policy relevance (D3) employment or housing policy objectives which will be picked up by attributes D1 and D2 should be excluded. The concern is not with possible correlation between schemes scores on different attributes, which is quite likely, but to avoid definitional redundancy.

Parsimony requires that, all else equal, the attribute set should be kept as small as possible, simply for ease of application. The attribute set used in COMPASS, has 32 attributes plus a separate capital cost assessment. The latter is kept separate in order to facilitate cost-effectiveness calculations, in view of the likely existence of capital budgeting constraints. 32 is an undesirably large number of attributes by the standards of normal multiattribute analyses. The reason for the size of the attribute set is the requirement that PATs act as production line techniques, assessing a wide range of projects without case-specific intervention. A fully comprehensive set of attributes must be specified and included in the structuring and parameterisation of the model, even though, in any one application, all attributes are unlikely to be needed. Consideration was given to the possibility of creating smaller models specific to different project types or cost bands, but the idea was rejected largely because direct comparability between the wide range of potential highway schemes was thought to be important. There is, for example, a view that small (but cost-effective) schemes receive less support in some existing planning procedures than larger, higher-profile possibilities. It is also worth noting that some of the apparent excessive size of the tree is accounted for by the specification of the attribute set as a form of checklist, to ensure that important impacts of schemes are not overlooked by inexperienced assessors. For example, all vehicle operating cost savings should arguably receive the same unit weighting and might in principle be aggregated under a single attribute. The size of the attribute set does undoubtedly pose some problems, as will be discussed in section 4.3.1 and 4.3.3, but seems inevitable, given PATs' applied and institutional context.

4.2 Assessing Impacts of Alternatives

In a conventional multiattribute application, the second phase of the analysis, once objectives and attributes have been identified, is to assess the impacts of the alternatives by specifying their "scores" on all the attribute scales. This information acts as an input to the third phase of the analysis. The fact that it is not available in the conventional sequence in developing a PAT leads to some difficulties in the scaling of scores, as will be described in the next sub-section.

4.3 Determining the Preferences of the Decision Makers

The objective in this part of a multiattribute analysis is essentially to elicit the decision makers' trade-offs between the specified attributes. In a normal decision analysis this would be achieved through direct, carefully structured interaction between the decision analyst and the decision maker(s). For our PAT, because there was no pre-identified single user group and because considerable experience exists in specifying trade-offs between at least some of the attributes involved, the research group itself specified the preference structure and estimated the trade-offs in the first instance. Four steps have to be undertaken

- determining the general preference structure

- assessing single-attribute value functions

- evaluating scaling constants

- checking for consistency

4.3.1 Determining the General Preference Structure

The effectiveness of multiattribute analysis depends on the ability first to address preferences on individual attributes (occasionally, but not in our case, on small groups of attributes) and then to combine that information into an overall preference model. Formally, we are looking to define a function, f(.) such that $V(x_1, \ldots\ldots, x_n) = f[v_1(x_1), \ldots, v_n(x_n)]$ where V(.) is the overall assessment of an alternative with attribute scores $x_1, \ldots, x_n$ and $v_j(x_j)$ are the individual, one-dimensional value functions on each attribute.

The first step in characterising f(.) is to identify the relevant preference structure, which is done by ascertaining whether or not certain preferential independence conditions between attributes hold.

Let $\{X\} = \{X_1, \ldots\ldots, X_n\}$ represent the set of attributes selected as characterising the alternatives under consideration and let $\{Y\} = \{X_1, \ldots\ldots X_s\}$ and $\{Z\} = \{X_{s+1}, \ldots\ldots X_n\}$ correspond to a pair of mutually exclusive and collectively exhaustive subsets of {X}. Then, following Keeney and Raiffa (1976), a set of attributes is said to exhibit mutual preferential independence if every subset, {Y}, of {X} is preferentially independent of its complementary subset, {Z}. In turn, {Y} is preferentially independent of {Z} if and only if

$$[\ (y', z') \geq (y'',z')\ ] \Rightarrow [(y',z) \geq (y'',z)\ ]$$

for all vectors y', y" and z of specific scores on the attributes. In other words, preferential independence requires that if attribute scores for the subset of attributes {Z} are common across two alternatives and $x' = (y',z')$ is preferred or is indifferent to $x'' = (y'', z')$, then changing the scores in the {Z} subset to a different but still common set (z") must not change the fact that the alternative with scores y' for {Y} will be preferred to the one with scores y".

Preferential independence is most readily established by taking advantage of a theorem of Gorman (1968) which states that if {U} and {V} are subsets of {X} which are preferentially independent of their respective complements, are such that {U} and {V} overlap (but neither is contained in the other) and are such that $\{U\} \cup \{V\} \neq \{X\}$ then:

(a) $\{U\} \cup \{V\}$

(b) $\{U\} \cap \{V\}$

(c) $\{U\} - \{V\}$ and $\{V\} - \{U\}$

(d) $\{\{U\} - \{V\}\} \cup \{\{V\} - \{U\}\}$

are each preferentially independent of their respective complements. This permits, for example, complete mutual preferential independence to be established simply by establishing preferential independence of each pair of attributes $\{X_i, X_{i+1}\}$ [i = 1, ...(n-1)]. Finally, if mutual preferential independence holds, then, given a number of other relatively straightforward and plausible conditions (French, 1988, p. 119/20) it is necessary and sufficient that the function f(.) defined at the beginning of section 4.3.1 is additive:

$$v(x_1, \ldots\ldots, x_n) = \sum_{j=1}^{n} w_j v_j(x_j)$$

Thus, if for our PAT, it is possible to establish mutual preferential independence, a major simplification of the modelling of priority assessment will be available.

The set of attributes derived for priority assessment (appendix 1) has some features which will help diminish the size of the task of checking for preferential independence. First, attributes B2.1 - B2.4, relating to operating cost savings are all measured in the same units, £K per year. They are named separately in the PAT list of attributes not because they are dimensionally different, but as a check on comprehensiveness and in order to form a checklist for users. Logically, however, a pound's worth of saving per year should be traded off against some other attribute identically irrespective of whether it arises through saving the time of lorry drivers or public transport users, say. If the decision-maker feels uncomfortable with this suggestion, then there is an issue which needs to be probed more deeply regarding either the value-tree structure or the attributes. It could be that the decision-maker has strong distributional views, e.g. favouring savings to individual members of the public rather than to businesses; or it may be that cost savings are acting implicitly as a proxy for other considerations, not articulated within the tree.

Despite these possibilities, we shall assume that, for the purposes of considering preferential independence, all cost savings can be treated as identical. Similarly, we do not distinguish between vehicle only and pedestrian accidents. In this way, the set of attributes to be considered is reduced from the 32 of appendix 1 to 26.

One convenient approach to analysing preferential independence is to consider the 25 pairs of attributes formed by the combination of attribute (A1.1.1/A1.2.1) - slight accidents - and each of the remaining attributes. For each pair, the following question is then considered:

> Should the rate at which the decision-maker would trade-off changes in the level of these two attributes be affected in any way by knowing the levels taken by any of the other attributes?

If the answer on each occasion is negative, then Gorman's theorem quoted earlier permits us to assert that there is mutual preferential independence within the attribute set.

To answer this set of questions is far from straightforward. Even with the reduction to 26 attributes, the problem is a very large one by normal decision analysis standards. It is difficult to focus on just two of the attributes and to consider how one would react towards different combinations of them as the other 24 potentially take different values. No formal checking with decision-makers was undertaken. However, our own introspection suggests that negative answers are defensible, at least as a reasonable approximation.

A second way to tackle the preferential independence question is hierarchically. For example, with only a minor alteration to the tree shown in figure 1, we can construct a set of seven aggregate variables:

A. Accidents
B. Highway characteristics
C. Travel time savings and delay during construction
D. Operating cost savings
E. Environment
F. Disruption to residents during construction
G. Planning and development.

The same style of pairwise preferential independence questions as was asked previously, may now be asked for (A,B) through to (A,G). Some individuals may find it more straightforward to think in terms of such a smaller set of aggregate attributes; others, of course, may find it harder, perhaps if it is difficult to envisage all the impacts that some of the aggregates capture. If preferential independence is established between the seven aggregate attributes, we can then go on to check for independence within the sets of lowest-level attributes from which A .... G were constructed. If independence is found within each set, then preferential independence has been established across all 26 attributes. Again, although no formal independence checks were carried out with decision-makers using the hierarchical approach, our own introspection suggests that it may reasonably be assumed. It is perhaps worth noting that, especially with the hierarchical form of check, if less than complete independence is established, it may still be possible to achieve a considerable simplification of the structure of the value function, v(.) (Keeney and Raiffa, 1976, p. 115/6). However, as far as the construction of a value function for our PAT was concerned, the analysis described above was considered adequate to justify moving ahead on the assumption of an additive model.

4.3.2 Assessing Single-attribute Value Functions

The next required step is to establish the functional form for each of the 32 $v_j(x_j)$ in the model. There are, in principle, many ways of doing this (Watson and Buede, 1987, p. 194). What was done in this case was strongly influenced by the working context of PATs.

The approach adopted was to identify, for each attribute, a worst (w) reasonable and best (b) reasonable score that highway projects in the cost range under consideration

(£25k to roughly £2m.) might achieve. These scores were then scaled such that $v_j(x_j^w) = 0$ and $v_j(x_j^b) = 1$. To evaluate intermediate levels of x_j, an interval scale was constructed:

$$v_j(x_j) = \frac{x_j - x_j^w}{x_j^b - x_j^w} \qquad \begin{cases} x_j^w \leq x_j \leq x_j^b & \text{if best connotes high } x_j \\ x_j^b \leq x_j \leq x_j^w & \text{if best connotes low } x_j \end{cases}$$

An implication of this scale is that equal increments in x_j occurring anywhere within the range for which it is defined imply equal value increments as well.

The decision to construct the scale in this way was made without any attempt to discover from decision-makers, with the aid of the established empirical techniques, whether a non-linear $v_j(x_j)$ might have been more appropriate in some cases. The judgement was made that, at least in the first instance, simplicity was the dominant requirement, because:

(a) experience suggests that non-specialists (not only local politicians but also engineering and other professionals) have a low tolerance threshold for any opacity in decision aids;

(b) linear $v_j(.)$ would ensure that the overall value function $V(.)$ behaved rather like the benefit calculation in a cost-benefit analysis, an evaluation method generally familiar to most people involved in transport planning;

(c) linear $v_j(.)$ made the question of dealing with individual x_j values outside the x_j^w to x_j^b range straightforward in a way which it could not be if the $v_j(.)$ were non-linear.

Even in the relatively simple working framework provided by the adoption of interval scales for the $v_j(.)$, however, there are a number of practical problems to be faced. First, there is the issue of determing x_j^b and x_j^w and the related question of selecting a scale of measurement for each x_j. An overall assessment of the level of achievement of many of the objectives considered by PATs depends on two factors, the extent of the change which occurs in a given variable and the number of people or vehicles affected by the change. Within the range of highway schemes addressed by PATs, wide variation in both occurs. x_j^b and x_j^w were thus defined as the products respectively of the best reasonable underlying attribute score and the maximum number of people/vehicles likely to be affected and the worst reasonable underlying attribute score and the minimum number of people/vehicles likely to be affected (zero in all cases). This general pattern was followed not only for objectively measurable variables such as vehicle operating cost savings (number of vehicles affected by the scheme per year x cost saving per vehicle) but also for a considerable number of other variables (highway characteristics; some environmental variables) where a subjective scale of assessment was used. Here x_j^w would be zero (no people/vehicles benefitting) and x_j^b would be the product of the maximum subjective improvement (10 points) multiplied by the maximum number of people/vehicles that might reasonably be affected by the scheme.

The use of subjective scales, although in our view inevitable in PAT construction, causes some difficulties. One of the reasons for choosing linear $v_j(.)$ was that x_j values outside the x_j^w to x_j^b levels built into the parameterisation of the multiattribute value model would inevitably occur from time to time. The typical PAT user is likely to be unwilling and/or unable to re-parameterise the model. With linear $v_j(.)$, objectively measured scales may be treated as open-ended without any detrimental effects. Analogously with cost-benefit analysis, there is no need to postulate any *maximum level of cost or benefit beyond which the validity of the* appraisal technique ceases. However, subjective scales, to be workable, must be closed, must have fixed minimum and maximum levels of achievement. The question then arises of what to do if a particular project performs on a subjective attribute at a level outside the range conceived when the model was first set up. The solution proposed to users of COMPASS is to score the project at the relevant extreme subjective assessment, but to "star" it as having special characteristics not fully accounted for by its numerical score, v(.). In practice, we envisage that such schemes will be rare and will be likely by their very nature to demand an element of special consideration that would almost certainly be afforded by existing administrative and political procedures.

A second and more pervasive problem associated with the use of subjectively assessed attribute scores in PATs is consistency. Even given the guidelines we have worked to on scheme capital cost, a very wide variety of projects is typically processed each year through local highway planning offices. Moreover different schemes (or even different aspects of the same scheme) are likely to be assessed by different individuals. Assuming that one person's subjective judgement is broadly comparable with that of anybody else using the PAT is clearly vital. It is essential also that each individual assesses each scheme against the full range of schemes covered by the PAT, and not just relative to schemes similar in type or cost scale to the one under consideration. There is no formal way of ensuring it. Calibrating the judgement of the individuals using the model must depend on departmental guidelines and shared practical experience. Given the number of subjectively assessed attributes in the model, it is a most important aspect of the functioning of any PAT system.

In defining the individual $v_j(.)$, elements of approximation and potential inconsistency are inevitable. This is the basic reason for choosing a simple linear functional form, even though some of the attributes are patently non-linear (e.g. changes in noise levels, measured in dBA). Nonetheless, should it transpire in practice that the interval scale assumption is too much at variance with decision-makers' judgements, replacing an individual $v_j(.)$ with a non-linear function should cause no insuperable difficulties, although it might then be necessary to widen the x_j^w to x_j^b range, if there have been many occurrences of x_j's outside the initially specified limits. Thereafter, occasional extreme x_j values would have to be "starred" in the same way that is suggested for extreme subjective scores.

4.3.3 Evaluating Scaling Constants

As with the determination of individual value functions, so with techniques for evaluating the scaling constants (attribute weights, w_j) there are many approaches available (see, e.g. Hobbs, 1980; Schoemaker and Waid, 1982; von Winterfeldt and Edwards, 1986). Although some of the methods may reasonably be set to one side, either on theoretical or practical grounds, there remain several, for any one of which a case can be made. This section sets out the broad categories within which the different methods fall, highlights some as serious contenders for application and explains how the initial default set of weights employed in COMPASS was derived.

Following Schoemaker and Waid, five broad categories of weight assessment technique may be identified: multiple regression (MR); analytic hierarchy (AH), direct trade-offs (DT); points allocation (PA) and unit weighting (UW). To use MR would require here a substantial data base of previous projects, specifying both attribute scores and some holistic index of overall project value, so that weights may be calibrated using standard regression procedures. Such a data base (especially the holistic evaluations) is unlikely to be available. Even in cases where it is, there can be problems in identifying a meaningful set of weights (Pearman, 1989). For both these reasons, the MR approach was not considered for our PAT. Similarly, PA (Metfessel allocation) was not applied. Although this method is straightforward (simply allocation a fixed number of "points" - say 100 - between the attributes according to their importance) there are doubts about the validity of the weight sets that result (Hobbs, 1980; Watson and Buede, 1987). It would also seem, _a priori_, difficult to apply the method consistently across a large attribute set.

A third of the approaches identified by Schoemaker and Waid was not pursued in detail for PAT application, the UW approach. UW requires that equal weight be given to all attributes, after they have been standardised, e.g, to equalise their means and standard deviations. As with MR, depending upon the type of standardisation undertaken, the required data set may not be available. Moreover, the validity of the arguments which suggest that equal weighting yields defensible multiattribute valuation models continues to be a matter of contention; insofar as a case exists for this approach, it seems unlikely that PAT models operate in the applied circumstances which justify the use of unit weighting (von Winterfeldt and Edwards, 1986, pp. 441-3). Nonetheless, given the simplicity of the UW method, some retrospective analysis comparing results derived from a non-UW PAT with those which would have resulted from a UW model would be interesting.

The two remaining approaches which seem to justify fuller consideration for use in a PAT are AH and DT. Either of these can, in turn, be applied in one of two ways. First they can be applied directly to the full set of 32 attributes. Alternatively they can be used hierarchically within the value tree structure. In the case of AH, a non-hierarchical application would require 496 pairwise comparisons of the relative importance of attributes, an impracticable task. Although techniques do exist for undertaking the AH calculations with "missing values" (Islei and Lockett, 1988; Harker, 1987), nonetheless, non-hierarchical AH was deemed not to justify serious consideration. Thus the three contenders are hierarchical AH, together with hierarchical and non-hierarchical DT.

The AH procedure in its standard form requires that the decision-maker should estimate all pairs w_i/w_j ($i < j$) of weight ratios. Since the weights are, by convention, normalised (e.g. such that $\Sigma w_j = 1$), in principle a completely consistent decision-maker would only have to estimate $(n - 1)$ ratios to fix all the required w_j. In practice, decision-makers are not consistent. The $\frac{1}{2}(n-1)(n-2)$ excess ratio estimates act as a form of consistency check in that, by one process or another, the AH technique derives a set of w_j estimates which are in some sense as consistent as possible with the full set of ratio estimates. In his own formulation, Saaty, the originator of AH, favoured the use of the principal eigenvector of the matrix formed by entering w_i/w_j in all cells (i,j) and 1 for all diagonal cells (i,i) (see e.g., Saaty, 1988). More recently, it has been argued (Barzilai et al, 1987) that more justifiable estimates of the weights come through calculating the geometric means

$$w^*_i = \left[\prod_{i=1}^{n} \frac{w_i}{w_j} \right]^{1/n}$$

following which the corresponding w_i are found through normalising by dividing by Σw_i^*.

The AH approach has excited a good volume both of praise and criticism (Zahedi, 1986). Praise centres on its acceptability to users, an important consideration for PAT implementations. The various forms of criticism ultimately relate to what the critics see as the lack of a convincing axiomatic foundation for the method, despite Saaty, 1986. Some concentrates on the way the AH technique is often implemented (e.g. failure to take into account units of measurement; Saaty's preference for limiting the w_i/w_j ratio to the range 1/9 to 9 via a verbal response scale). Neither of these need apply to the case of using AH to estimate weights for a PAT. Another and more fundamental area of criticism revolves around the various different ways which have been suggested for computing the best estimate of the w_j, once the pairwise comparison ratio matrix is known. This point is examined further in Appendix 2. For the present, a reasonable summary seems to be that AH provides a readily implementable way of undertaking the weighting stage of multiattribute value function construction. The expressed doubts about its theoretical validity should be seen in the light of the accuracy of the data processed by PAT's and their status as decision aids, not decision prescribers.

The alternative set of weighting procedures available is DT. Within this set, there are many possibilities in terms of detailed implementation. Frequently, the starting point is a ranking of the attribute weights, derived with particular attention to the attributes' units of measurement. Ideally, this is obtained by asking the decision-maker to rank the changes from best level to worst level in each attribute while all other attributes are held constant, e.g. at their worst level. The ranking of the w_j, once obtained, can be processed in a number of ways to obtain normalised weights on a ratio scale. Some (like the rank reciprocal rule, $w_j = 1/r_j\ [\Sigma\ 1/r_j]$, where r_j is the rank of attribute j) are simple rules of thumb applied without any further analysis specific to the individual problem. Alternatively, methods like Edwards' ratio technique (Edwards, 1977) require further, application-specific inputs, in the ratio technique case, successive estimates up the ranking of w_j/w^{worst}, (with w^{worst} being fixed at an arbitrary figure such as 10 points) followed by normalisation to ensure that the weights sum to unity.

A more sophisticated approach works not directly with the weights themselves, but in the following way. Suppose the attributes are relabelled such that x_1 corresponds to the first ranked attribute, x_2 to the second and so on. x_1 may then be used as a numeraire in a succession of questions, the first of which is : at what value of x_1,

x_1^i, would you be indifferent between an alternative $x_1 = (x_1^i, x_2^w, \dots x_n^w)$ and $x = (x_1^w, x_2^b, x_3^w, \dots\dots x_n^w)$? Once the point of indifference is established, it is clear that

$$w_1 \; v_1(x_1^i) = w_2$$

Repeating this process a further (n − 2) times and adding the normalisation condition $\Sigma w_j = 1$ yields a set of equations through which the values of all the w_j may be computed.

At least in the present state of the art, as Schoemaker and Waid (1982) have noted, choosing a weighting technique is itself a multicriteria choice problem, involving considerations such as ease of use, mean performance, axiomatic justification and trustworthiness. It is important to bear such factors as these in mind and also to exploit information and/or respond to constraints arising from the particular application concerned. For example, as far as our PAT is concerned, a single-sweep application of any of the DT techniques, involving a minimum of 31 direct trade-off calculations, would seem a particularly demanding task. Thus, unless it is decided to exploit particular characteristics of the PAT problem to simplify the process, weight assessment by either AH or DT needs to be approached as a hierarchical problem.

The hierarchical calculation of weights may in principle be undertaken either top-down or bottom-up through the value tree. However, since the branch descriptors used at the aggregate level in many trees (including ours) do not have any natural units of measurement associated with them, it is often necessary to proceed bottom-up. The method is as follows. Using whatever AH or DT approach is preferred, normalised weights are first computed within each cluster of lowest level attributes, e.g. A1.1.1 through to A1.2.3 in Table 1. Once this has been done, a single representative attribute is chosen to represent each cluster and a calculation of normalised weights is again undertaken between each of the representatives. If there are more than two levels in the tree hierarchy, then the process is repeated again as many times as is required to reach the top of the tree, each time selecting a single lowest-level attribute to represent the sections of the tree that are being compared. Appendix 2 illustrates the application of the AH technique to the value tree shown in Figure 1 to derive the required weights.

There is some evidence (Stillwell et al, 1987) which suggests that the results obtained by deriving weights hierarchically, rather than flat across all (in this case, 32) lowest-level attributes, exhibit greater "steepness" – that is, numerical differentiation between attributes. A direct check of this observation is not possible on the basis of our model, because of the perceived impracticability of assessing all 32 weights in a single sweep. Instead the variation of the DT technique which we employed initially to derive a set of "default" attribute weights rested upon a more pragmatic approach, exploiting and responding to certain characteristics of the PAT problem, but in principle being a variant of the "pricing out" procedure (Keeney and Raiffa, 1976, pp. 125–9).

A number of the attributes in our PAT are assessed directly in money terms; in addition, the Department of Transport, through procedures such as COBA, (Department of Transport, 1981) has traditionally provided monetary estimates of a number of the other attributes in the value tree. Although the money values attached to such items as time–savings and accident avoidance are contentious, nor necessarily the values to which any particular local authority would wish to adhere in its decision making, they nonetheless provide a starting point for weight formulation which is helpful because of its familiarity to potential users. Thus, using operating cost savings as a natural, money–based numeraire, the relative weight of unit changes in variables such as accident and time savings were estimated. Weights for those attributes which could not be handled in this way (the environmental attributes and those assessed on a subjective scale) were established by trying to estimate money values for the consequence of moving one individual or some similar identifiable unit, from the worst likely to the best likely extremes of the scoring scale. All the relativities so assessed then had to be re–scaled for the assumptions described earlier about the levels of x_j^b and x_j^w and about the maximum number of units likely to be affected, before the weights were finally normalised to sum to one. It was these weights which were used as the initial set of default weights in the computer implementation, COMPASS.

Thus the position overall about the assessment of scaling constants is that a number of acceptable techniques are available, with no one method exhibiting substantial general advantages over the others (Schoemaker and Waid, 1982). At present COMPASS provides users with a default set of weights derived using an ad hoc variant of the pricing out procedure. However, weight derivation using DT or AH methods is possible and would have the advantage of being somewhat less influenced by the conventional relative values operating at present. An important question in weight derivation is the extent to which it is desirable to gather redundant information to provide consistency checks on weight estimation and the way in which all the information elicited should be combined to yield weight estimates. It is hoped to explore this issue, along with the acceptability to users of alternative approaches to weight derivation in co–operation with local authorities using the prototype version of

COMPASS.

4.3.4 Checking for Consistency

The standard decision analysis sequence requires, as the final stage of the process of determining decision makers' preferences, that the initial assessments be checked for consistency. In the context of our PAT, this process took the form of discussion between the research team members of their independent attempts to define the unidimensional value functions, $v_j(.)$ and to compute the w_j, and the application of COMPASS to a series of six trial projects. The main lesson which emerged from this exercise was the great importance which attaches to the decision maker being clear in his/her own mind about the units of measurement and the scale minima and maxima, x_j^w and x_j^b, when computing the w_j.

4.4 Evaluation of Alternatives

The three previous stages (problem structuring; assessing impacts of alternatives; determining the preferences of decision makers) have structured and parameterised a multiattribute value function for the task of prioritising local authorities' highway schemes. The application of the model will associate an aggregate score, V(.), ($0 \leqslant V(.) \leqslant 1$) with each candidate scheme, such that the higher is V(.), the more preferred is the scheme. It is important to acknowledge, however, that there are at least three sources of potential error in a model of this type. First, there may be data errors relating to individual projects. Secondly, there may be errors in estimating the w_j. Thirdly, there may be errors in the structure of the model itself. Although no set of checks or other procedures can guarantee to eliminate all such errors, steps can be taken to try to diminish their consequences.

As far as the first two of the sources of potential error are concerned, the principal defence is sensitivity analysis. Procedures must be provided to facilitate checking how scheme ranking might be affected by changing project scores and/or weights. Such checks are important not simply in a technical sense, but also psychologically. The ability to demonstrate to decision makers the extent to which choices may or may not be robust to changes in input values often has a substantial influence on the acceptability in practice of a model's recommendations. Sensitivity analysis in multiattribute modelling is, however, much more art than science; there is no single set procedure which can be specified.

At present, the sensitivity analysis provided in COMPASS is quite basic. As far as sensitivity to weight changes is concerned, two types of analysis are possible. The user may change either the weight on one lowest-level attribute or the aggregate weight attached to any one of the four major attribute sub-divisions in the value-tree hierarchy, safety, traffic, environment or planning. In either case, the remaining weights are renormalised to ensure that they sum to one, keeping all the other attributes weighted in the same proportion to each other as they were initially. Project rankings may then be directly compared using the old and new weight sets. For sensitivity to changes in attribute scores, the facility exists to amend the project scores and re-analyse the amended project or projects in order to assess the effect on the final ranking of the changes which have been introduced.

Sensitivity testing in the initial version of COMPASS[1] has been kept straightforward for a number of reasons

- it is not yet clear whether potential users are likely to require any more sophisticated sensitivity test facilities, or, if so, which kinds;

- since the initial version of COMPASS works in conjunction with the Lotus 1-2-3 spreadsheet, it is not possible to program in an efficient way all the types of sensitivity test that are potentially useful, especially those that require good graphical facilities;

- there seem to be few clear guidelines from the decision analysis literature as to what are likely to be effective forms of sensitivity analysis in circumstances such as those COMPASS is modelling.

Finally, it should be noted that the COMPASS user is presented with a choice of ranking criteria (aggregate score V_i; V_i to capital cost ratio; V_i to capital cost minus construction grants ratio; V_i to capital cost minus construction grants minus annuitised change in maintenance cost ratio). Different sensitivity analysis procedures might well be appropriate, depending upon the chosen criterion.

The focus of much of the debate about sensitivity testing in multiattribute decision analysis has been the flat maximum principle. This suggests that linear evaluation models are remarkably robust to changes in weights and project attribute scores. If this were true, it would imply that the key aspect of any PAT was the identification of the appropriate set of attributes and that, once the correct attribute set was chosen, sensitivity testing on scores or weights derived on any reasonable basis would be unlikely to show much variation in calculated project values, V, as the inputs were adjusted. Most decision analysts agree that the problem structuring phase of a decision analysis is of great importance. There is much less unanimity about the flat maximum principle.

The literature suggests that flat maxima are most likely when, within the set of alternatives, the attribute scores are positively correlated and when the number of attributes is relatively small (von Winterfeldt and Edwards, 1986, p.443). The latter is certainly not the case in our PAT, and the former is questionable. The likelihood of a flat maximum is also increased if dominated alternatives (those that could never be optimal under any set of w_j) are first removed. It should be noted, however, that the presence of a flat maximum defined in this way does not necessarily prohibit changes in which project ranks first out of a set.

At present, COMPASS undertakes no checks for dominance. Indeed, checking for dominance would seem to be far from straightforward. PAT's are concerned, typically, not with identifying a single "best" project, but with picking "the best k from n projects". However $^k/_n$ selection dominance analysis is complicated by the fact that the dominance structure will change as the best projects are creamed off and treated as firmly accepted. Some progress has been made with this problem (John et al., 1980). However, to complicate matters further, PAT selection is not simply $^k/_n$, but $^k/_n$ with a capital cost budget constraint. Where a capital cost constraint exists, the most practicable cost-benefit analysis procedure is to select projects according to the ranking of their NPV/capital cost ratios until the budget is exhausted, although even this is an approximation to a truely optimal selection procedure (Pearce and Nash, 1981, pp. 46/7.) By analogy, the relevant consideration in a PAT's analysis is the weighted score (V_i) to capital cost ratio. If, as some people suspect, there is a tendency for cost-effective small schemes to receive less favourable treatment than they should, sensitivity testing on V_i alone would seem to be of value largely in the initial stages of using a package like COMPASS, when decision makers are starting to come to terms with how weight changes affect the relative standing of schemes. For more detailed analysis, the most effective procedure at present would seem to be to test the sensitivity of the rank ordering of projects' V_i capital cost ratio (with capital cost defined in whichever of the three ways the user chooses) in the face of carefully chosen alterations in the weights attached to attributes or groups of attributes whose relative importance is least confidently understood.

The third form of sensitivity test which should in principle be attempted at this stage is to assess the extent to which the chosen model of preference structure still seems to be the correct one. Again, because we do not have direct access to the final user, we cannot make this assessment in the conventional way, by assessing the decision maker's reaction to the model's performance. Instead, we have to rely on more general evidence which is available about model structure. There are two matters of principle concern. One is the adequacy of a linear additive model to represent the preference structure; the second concerns the choice of attributes within the linear structure.

The multiattribute model we have constructed is linear both in its overall structure and in that the individual value functions, $v_j(.)$, for each attribute are linear. This latter assumption has not been fully tested. There are clearly some scales, such as noise measurement, where it will be important to consider how users respond to the linear scale and replace it, if necessary, with an appropriate non-linear $v_j(.)$. It is also the case that all the attribute scales were created on the assumption that the highway schemes under consideration would improve system performance as measured by the attribute concerned, or at least make it no worse. Because of the linearity of the $v_j(.)$, schemes with negative scores cause no problem in undertaking the computations. However, we have not checked explicitly whether the weight decision makers wish to give to a deterioration in performance is indeed simply the negative of the weight they would give to an equivalently sized improvement. If not, then some amendment of the $v_j(.)$ would be needed.

As discussed earlier, only limited checks were undertaken as to whether the structure of preferences justified assuming that the multiattribute model as a whole should be linear. In a conventional decision analysis, the acceptability of the linear model would be tested through experience of its use by the decision maker. At this stage, for our PAT, we have to rely instead on the correctness of the initial judgement about the existence of preferential independence and on the view of many practising decision analysts (e.g., Dyer and Larsen, 1985) that linear additive models provide in practice a very effective approximation to true underlying model structures for decision making, especially if the objective is to identify a number of promising projects, not just the single "best".

A related issue of some significance is the identification and definition of attributes within the linear model. Whether particular types of highway scheme impact have been omitted is something which is likely only to emerge as the PAT is used. Two other questions, however can be addressed more immediately. One is the likely effect in general of omitted variables. Here there is some ambiguity, but, if sensitivity to missing attributes is measured by loss of value of the chosen alternative (Barron and Kleinmuntz, 1986) rather than correlation across the evaluations of all alternatives (Kleinmuntz, 1983), it seems that omitted variables can, at least in some circumstances, be important. The second question is the level of detail in which different areas of impact are assessed - in the sense of the number of attributes allocated to each area. Recent work (Weber et al, 1987) suggests that parts of a value tree which are represented in more detail will be systematically over-weighted. This, not altogether counter-intuitive finding ties in with a finding in our initial survey of PAT's (Simon, 1986a) that some PAT's used no attribute weighting at all and must therefore have relied on implicit weighting through the identification of different measurement scales and/or number of attributes in given areas to impose a weighting of components on overall scheme assessment.

One of the reasons for creating a hierarchically structured value tree was to help keep a check on this type of potential bias. In the absence of knowledge of the 'true' model, vigilance is perhaps the best protection. The four major components of our model are represented respectively by 10,11,8 and 3 lowest-level attributes, suggesting a priori that it is planning and development considerations which might end up under-valued. In the longer term, some type of check is possible. Unfortunately, it cannot take the simple form of aggregating the lowest-level weights in each of the sections of the tree (giving for the default set of weights respectively Safety = 0.452; Traffic = 0.025; Environment = 0.265 and Planning = 0.258). This is because the numerical weights reflect not simply the relative importance of the attributes, but also the chosen scales of measurement for the individual attributes. If the scales are changed (e.g; x_j^b and/or x_j^w) the w_j will change also. Some progress can, however, be made once a few (not untypical) schemes have been processed through the system. For each scheme, the sum

$$\sum_{j \epsilon S} w_j \; v_j(x_{ij})$$

(S = A, B, C, D - in the terminology of column 1 of appendix 1) will give the contribution of attributes in the four major sub-divisions of the value hierarchy to its overall assessment, $V_i(.)$. One of two procedures may then be chosen. If the $\Sigma\, w_j\, v_j(x_{ij})$ are summed across schemes and then put on a percentage basis, the result expresses the average contribution of each of the four major impact types to schemes undertaken by the authority. Alternatively, normalisation may first be undertaken across each individual scheme, followed by summation and a second normalisation across the four headings. In this case we are calculating the contribution of each of the four impact areas to a "typical" scheme undertaken by the authority. Schemes with low $\Sigma\ w_j\ v_j(x_{ij})$ (and therefore, presumably, low cost) are weighted equally with all other schemes. Table 3 suggests that, within the small set of schemes examined, smaller schemes have a substantially higher safety orientation and are less effective on traffic and planning issues.

	Attribute Heading	Scheme 1	2	3	4	5	6
A	SAFETY	0.034	0.063	0.084	0.082	0.116	0.238
B	TRAFFIC	0.003	0.031	0.009	0.004	0.027	0.410
C	ENVIRONMENT	0.000	0.001	0.008	0.000	0.000	0.083
D	PLANNING	0.009	0.017	0.017	0.009	0.017	0.427
	TOTAL	0.046	0.112	0.118	0.095	0.160	1.158

Table 1 $\sum_{j \in s} w_j\ v\ (x_{ij})$ for the Six Trial Schemes

	Attribute Heading	Scheme 1	2	3	4	5	6
A	SAFETY	0.739	0.563	0.712	0.863	0.725	0.206
B	TRAFFIC	0.065	0.277	0.076	0.042	0.169	0.354
C	ENVIRONMENT	0.000	0.009	0.068	0.000	0.000	0.072
D	PLANNING	0.196	0.152	0.144	0.095	0.106	0.369
	TOTAL	1.000	1.001*	1.000	1.000	1.000	1.001*

(* rounding error)

Table 2 Scores with Each Scheme's Scores Normalised to 1

	Attribute Heading	Schemes Weighted Differently	Schemes Weighted Equally
A	SAFETY	36.5	63.4
B	TRAFFIC	28.7	16.4
C	ENVIRONMENT	5.4	2.5
D	PLANNING	29.4	17.7

Table 3 Average Contribution (%) of the Four Attributes to the Effectiveness of Schemes as a Whole

5. DISCUSSION

The purpose of this final section is to draw attention to a number of points relating to the use and/or further development of COMPASS and the multiattribute evaluation model on which it is based.

One of the key influences underlying the form in which COMPASS has been developed is the need to ensure that the appraisal process itself is cost-effective. Within the cost range of schemes which COMPASS is designed to analyse, cheaper schemes may well on occasions not justify the time and manpower requirements of appraisal against all 32 lowest-level attributes. Alternatively, it may be desired to run a simple, quick evaluation on schemes which are at an early stage in the design process. In these circumstances, the hierarchical structure of the value tree provides two ways in which appraisal can take place without explicitly addressing all 32 attributes. Each is a form of "retreating up the tree", to permit an assessment of the combined effect of a group of lowest-level attributes to be introduced as a single assessment at a point further up the tree hierarchy.

The first approach may be considered as a representative impact argument. The decision maker selects the single lowest-level attribute which he/she regards as best typifying the performance of the project as a whole with respect to the set of impacts from which the representative impact is taken. For example, the change in the number of slight vehicle-only personal injury accidents might be thought to parallel a scheme's performance with regard to changes in all accident numbers. The the contribution of changes in accident numbers to the scheme's overall evaluation is approximated by the score $v_j(x_j)$ on the chosen representative attribute multiplied by the total weight, Σw_j, associated with all impacts in the group.

In practical terms, this may be done within COMPASS in two ways. If only one project is being evaluated at the current time, a new weights "file" can be constructed, with all accident weights set to zero, except slight vehicle-only p.i.a.'s, whose weight is set to the sum of all accident weights. All accident number changes are set to zero, except for vehicle-only p.i.a.'s, where the estimated numbers (before and after) are entered in the usual way. The rest of the computation of the scheme's aggregate score is automatic. The difficulty with this procedure is that it breaks down if any other schemes, evaluated with the full set of weights or with different representative attributes employed, must be examined simultaneously, e.g. for sensitivity analysis. If this is required, then all schemes must employ the same weight set, and the effect of using a representative attribute must be achieved by adjusting scheme scores rather than scheme weights. At present, this must be done manually, using a simple conversion procedure (see appendix 3) which enables a raw (unstandardised) score to be computed for any attribute, such that the normalised score will be the same as the normalised score for some reference attribute. For any group of attributes (e.g., accidents) a raw score is entered for all lowest-level attributes such as to ensure that the normalised score $v_j(.)$ is the same for all attributes within the group. Although this is a more cumbersome process in terms of data input, it then permits sensitivity analysis to be undertaken automatically, using all the standard procedures available in COMPASS.

A second approach to economising on data input is available if the decision maker cannot identify a single lowest-level attribute which is adequately representative of the group for which an aggregate assessment is needed. With this approach, the decision maker chooses the group of attributes for which he/she wishes to avoid the necessity of a detailed assessment (e.g. all category C variables, Environment) and then makes a single subjective assessment in a 0-10 scale of how the scheme concerned performs in terms of environmental impacts. As with all subjective scaling in PATs it is important that the decision maker bears in mind that the reference group for comparison is all that local authority's schemes which might be evaluated by the PAT and not just those similar in scale, type or cost to the one under consideration. Once the subjective assessment has been made, a simple table look-up procedure (appendix 4) enables equivalent scores to be entered for all lowest-level attributes within the relevant group, and COMPASS analysis can proceed in the normal way.

A second area in which decision-makers may feel unhappy with the degree of precision which a multiattribute model of the type derived for COMPASS demands concerns the specification of the w_j. It can be argued that one of the disadvantages of the multiattribute value/utility theory modelling paradigm is that it leads to models overspecified relative to what is needed to make the required decisions. (Vincke, 1986). Similar sentiments are expressed by Phillips (1984). In doing so, it faces the decision maker with a daunting array of judgements to be made in structuring and calibrating the model. The quality (and hence reliability) of information elicited may not match the quantity. A natural way to respond to these concerns is to examine how the output of the model is affected if the w_j are not specified as single, fixed numerical values, but are allowed some flexibility - either through not demanding a single figure estimate initially, or by permitting some variation about the estimate of w_j, once it is made.

A number of writers (e.g., Kofler et al, 1984; Hazen, 1986; Scherer et al, 1987;

Weber, 1987) have considered choice problems with incomplete information. Much of this work is oriented towards application, but there seems to be relatively little published evidence about practical experience with different methods, especially with large-scale real-life multicriteria problems. One difficulty, which is encountered straightaway in partial information choice models is potential ambiguity about the choice criterion on which the formal ranking of alternatives will ultimately depend. Some progress can, of course, still be made using dominance and related ideas to identify potentially optimal and definitely non-optimal alternatives, but, with the w_j variable, the power of such methods is likely to be limited. Thereafter, a number of criteria have been suggested, all within the context of selecting a single "optimal" alternative. Examples include: maximising the minimum achievable weighted score; choosing the project which is preferred to all alternatives in the largest hypervolume consistent with the uncertainty about the weights; choosing the project with the maximum weighted score at a single representative point (e.g., the median point) within the n-dimensional space consistent with the uncertainty about the weights. Each has something to recommend it, but none has a sufficiently firm axiomatic basis that one can feel comfortable with choice based on just one alone.

It has, however, throughout been one of the foundations of COMPASS that it could and should act only as a decision support device. Within such a framework, the inability to identify a single choice mechanism is of less concern. Moreover, since nearly all the information about non-specific w_j is in the form of linear constraints, a desirable extension to COMPASS would be:

(1) to feed information about the (linear) constraints on the w_j and about project scores into a separate analysis module;

(2) to identify schemes which, under any weighting system within the prescribed bounds must be definitely included or definitely rejected from any short-list;

(3) to rank non-excluded schemes using each of the three criteria previously described.

A good deal of this analysis could be undertaken with standard linear programming techniques. The key question and one which is essentially empirical, is whether the set of linear restrictions on the w_j are tight enough to permit the alternative scheme rankings which emerge to have some substance in relation to the relative merits of the schemes themselves, rather than simply implying that a wide range of valuations for individual schemes is consistent with the given information on the w_j.

One final development from which COMPASS might benefit is a means of bringing within the scope of the formal analysis an assessment of problem severity. It is clear that some local authorities give substantial weight to this question, more or less formally, in their priority assessments. Particularly from the point of view of local political influence, being seen to make some attempt at solving a severe problem may be better regarded than making what is (technically or economically) a much more effective investment affecting an issue which does not have a high public profile.

The question of problem severity assessment raises a number of interesting issues. One is the broad philosophical problem of the extent to which the allocation of public funds should be influenced by some people's perceptions of a problem if there exists evidence to suggest that taking account of such perceptions leads to a demonstrably inefficient allocation of scarce resources. But if, for whatever reason, problem severity is regarded as something which needs to be taken into account, the question then is, how?

One way forward would be to construct a multicriteria severity index, broadly on the same principles as the effectiveness index V(.) which underlies COMPASS. Two-dimensional plots of severity against effectiveness could then easily be created for the decision makers, dominating and dominated schemes could be identified and decision makers generally be made aware of the opportunity costs of choosing to attack high-profile problems at the expense of low-profile solutions.

If such an approach is followed, two questions need to be addressed in constructing the index. The first is a problem structuring question. Should it be, as a matter of principle, that the attribute value structure for severity assessment is identical to that for effectiveness assessment? There is certainly some appeal in the argument that says that severity ought to be assessed in the same general dimensions as effectiveness, but practicality, if nothing else, suggests that the value tree for severity measurement, even if it has the same general structure as Figure 1, will be different. It may, in general, be less dense and have different attribute scales and weights.

The reason is that, if COMPASS and similar PAT's are truely decision support systems, they must respond to the thinking of their users. Almost certainly, in most decision makers' eyes, problem severity will be construed in terms of a limited number of variables. The density of the effectiveness value tree stems largely from its role as a checklist for assessments. It is not necessarily the case that severity as decision makers would wish to take it into account, would naturally be envisaged at that level of detail. Nonetheless, and bearing this point in mind, an initial appraisal of the attributes used for effectiveness assessment in COMPASS suggests that the majority would be implicit as attributes in the severity assessments decision makers might wish. Exceptions might be C5 (a direct consequence of scheme implementation) and D1–D3 (which are difficult to assess in a site-specific way). Since there are some arguments of principle in favour of a common assessment basis and since economy of data-gathering favours this also, it might well be desirable at least to start from the position that the relevant attributes for severity assessment are the 28 remaining from the effectiveness value structure after the four detailed above are removed.

As well as the problem-structuring issue just discussed, severity assessment poses a second important problem, which is a measurement one. Effectiveness of achievement relative to any one attribute is measured in COMPASS as a predicted difference (with scheme minus without scheme), scaled on to the 0–1 line. Problem severity is not, however, amenable to measurement in this way. Severity must be measured relative to some expectation or standard; it is probably also a ratio like:

$$\frac{\text{Without Scheme} - \text{Ideal}}{\text{Without Scheme}}$$

or

$$\frac{\text{Without Scheme}}{\text{Ideal}}$$

since the perception of the severity of the problem is almost certainly made relative to the specific circumstances of the scheme concerned, rather than as a difference from a general (non scheme specific) ideal. That is, decision makers' views on severity are something like "this is a dangerous crossing" (as crossings of this type go), rather than "this crossing has a large number of accidents relative to an ideal of no accidents at all". Thus, in any attempt to add a severity aspect to COMPASS, new scalings (and perforce new weightings) of the measured attributes will be necessary (by no means a trivial task), even if broadly the same set of attribute labels can be retained.

Incorporating the above points represent long-term goals in the development of formalised decision support for local authority highway investment decision making. In summary, the present position is that COMPASS provides a computer-based decision aid for this problem which:

- sets out a comprehensive list of impacts to be taken into account in assessing highway investment schemes of the scale relevant to this aspect of local authority decision making
- suggests cost-effective ways of measuring the impacts on appropriate attribute scales
- provides a straightforward linear additive multicriteria model to aggregate the attribute scores with a single overall score
- provides simple graphical output and sensitivity analysis capabilities to assist in scheme ranking
- allows the decision makers the option to specify their own weights within the multicriteria model
- permits small or partially-specified schemes to be evaluated in less than complete detail, but on a basis providing comparability with fully specified schemes

It is hoped that COMPASS provides a decision support environment tailored to the needs of those officers and local politicians who need a means of setting out an initial broad ranking of highway investment proposals and consistent with the time and resource constraints under which such assessments are made in practice. COMPASS is presently under test with local authority highway departments; further developments in its structure and capabilities are anticipated.

ACKNOWLEDGEMENT

Those of us who have worked on aspects of COMPASS (Peter Mackie, Tony May, Alan Pearman, Danny Ponton, David Simon and Ricardo Subiabre) would like to express our appreciation to the ESRC for financial support and to the officers of the many local authorities who have given us their co-operation.

FOOTNOTE

1. A second version of COMPASS has now been created (Subiabre, 1989). It is based on a data-base program (Clipper) rather than on a spreadsheet. Analysis times are considerably shortened and there is much greater flexibility to specify new attributes, and to re-define and re-scale the default set of attributes.

REFERENCES

Adelman, L., Sticha P J and Donnell M L (1986) An experimental investigation of the relative effectiveness of two techniques for structuring multiattribute hierarchies, *Organisational Behaviour and Human Decision Processes*, Vol. 37, pp. 188–96.

Barron F H and Kleinmuntz D N (1986) Sensitivity in value loss of linear multiattribute value models to attribute completeness, in B Brehmer, H Jungermann, P Lourens and G Seven (eds.) *New Directions in Research on Decision Making*, Elsevier Science Publishers, Amsterdam.

Barzilai J, Cook W D and Golany B (1987) Consistent weights for judgements matrices of the relative importance of alternatives, *Operations Research Letters*, Vol. 6, pp. 131–4.

Brownlow S A and Watson S R (1987) Structuring multi-attribute value hierarchies, *Journal of the Operational Research Society*, Vol.38, pp. 309–17.

Buede D M (1986) Structuring value attributes, *Interfaces*, Vol. 16, No. 2 pp. 52–62.

Crawford, G.B (1987) The geometric mean procedure for estimating the scale of a judgement matrix, *Mathematical Modelling*, vol. 9, pp. 327–34.

Department of Transport (1981) *COBA User Manual*, Department of Transport, London

Dyer J S and Larsen J B (1985) Using multiple objectives to approximate normative models, *Annals of Operations Research*, Vol. 2, pp. 39–58.

Dyer J S and Sarin R K (1979) Measurable multi-attribute value functions, *Operations Research*, Vol. 28, pp. 810–22.

Edwards W (1977) How to use multiattribute utility measurement for social decision making, *IEEE Transactions on Systems, Man and Cybernetics*, Vol. SMC-7, pp. 326–40.

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

Gorman, W M (1968) The structure of utility functions, *Review of Economic Studies*, Vol. 25, pp. 367–90.

Harker, P.T (1987) Shortening the comparison process in the AHP, *Mathematical Modelling*, vol. 8, pp. 139–141.

Harness, M D and Sinha K C (1983) *Priority Setting of Highway Improvement Projects*, Joint Highway Research Project Paper 83–9, School of Civil Engineering, Purdue University, West Lafayette, Indiana.

Hazen, G B (1986) Partial information, dominance and potential optimality in multiattribute utility theory, *Operations Research*, Vol. 34, pp. 296–310.

Himanen, V (1987) *Transportation Investment Evaluation: Basic Problems and Concepts*, Research Note 756, Technical Research Centre of Finland, Espoo.

Hobbs, B F (1980) A comparison of weighting methods in power plant siting, *Decision Sciences*, Vol. 11, pp. 225–37.

Hwang, C L and Yoon K (1981) *Multiple Attribute Decision Making: A State-of-the-Art Survey*, Springer–Verlag, Berlin.

Islei, G and Lockett A G (1988) Judgemental modelling based on geometric least square, *European Journal of Operational Research*, Vol. 36, pp. 27–35.

Janes, F R (1987) *Setting Priorities for Highway Schemes in Hertfordshire Using Interpretive Structural Modelling*, Technical Report DSS/FRJ/246, Department of Systems Science, City University.

John, R S, Edwards W, von Winterfeldt D and Barron F H (1980) *Equal Weights, Flat Maxima and Trivial Decisions*, Research Report 80–2, Social Science Research Institute, University of Southern California, Los Angeles.

Keeney, R L (1981) Measurement scales for quantifying attributes, *Behavioural Science*, Vol. 26, pp. 29–36.

Keeney, R L and Raiffa H (1976) *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*, J Wiley, New York.

Kleinmuntz, D (1983) *The Sensitivity of Linear Multiattribute Utility Models to Attribute Formulation*, Working Paper 82/3–4–27, Department of Management, University of Texas at Austin.

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

Leleur, S (1985) *Project Evaluation – Survey and Trends of Project Evaluation Approaches in Transportation Planning*, Report No. 45, Institute of Roads, Transport and Town Planning, Technical University of Denmark, Lyngby.

Mackie, P J, May, A D, Pearman, A D and Simon, D (1988) A Computer-Aided Priority Assessment Technique for Local Authority Transport Projects, in *Highway Appraisal and Design*, Proceedings of the PTRC Summer Annual Meeting, Seminar E, pp. 161-9.

Pearce, D W and Nash C A (1981) *The Social Appraisal of Projects*, Macmillan, London.

Pearman, A D (1989) Multicriteria analysis and public sector decision making, in S. Conrad (ed.) *Assignments in Applied Statistics*, John Wiley, Chichester.

Phillips, L D (1984) A theory of requisite decision models, *Acta Psychologica*, Vol. 56, pp. 29-48.

Roy, B (1985) *Methodologie Multicritere d'Aide a la Decision*, Economica, Paris.

Saaty, T L (1986) Axiomatic foundation of the analytic hierarchy process, *Management Science*, Vol. 32, pp. 841-55.

Saaty, T L (1988) *Multicriteria Decision Making: the Analytic Hierarchy Process* (2nd edition), RWS Publications, Pittsburgh.

Scherer, W T, Stewart B S, Sykes E A and White C C (1987) A new interpretation of alternative pairwise comparisons for a generalisation of SMART, *IEEE Transactions on Systems, Man and Cybernetics*, Vol. SMC-17, pp. 666-70.

Schoemaker, P J H and Waid C C (1982) An experimental comparison of different approaches to determining weights in additive utility models, *Management Science*, Vol. 28, pp. 182-96.

Simon, D (1986a) *Transport Priority Assessment Techniques developed by British Local Authorities: a Synopsis and Review*, Working Paper 230, Institute for Transport Studies, University of Leeds.

Simon, D (1986b) *Priority Assessment Techniques Project Phase 2: Detailed Appraisal and Testing of a Sample of Methods*, Technical Note 195, Institute for Transport Studies, University of Leeds.

Simon, D (1987) *Comparison of Scheme Rankings Using Six Priority Assessment Techniques*, Working paper 237, Institute for Transport Studies, University of Leeds.

Simon, D, Mackie P J, May A D and Pearman A D (1987) Assessing Local Authority Highway Priorities, in *Highway Appraisal and design*, Proceedings of the PTRC Summer Annual Meeting, Seminar E, pp. 255-66.

Stillwell, W G, von Winterfeldt, D and John, R S (1987) Comparing hierarchical and non-hierarchical weighting methods for eliciting multiattribute value models, *Management Science*, vol. 33, pp. 442-50.

Subiabre R. (1989) *COMPASS: A Flexible Priority Assessment Technique for Local Authority Transport Projects* M.A. dissertation, Institute for Transport Studies, University of Leeds.

TRB (1984) *Evaluation Criteria and Priority Setting for State Highway Programs*, NCHRP Synthesis of Highway Practice No. 84, TRB, Washington.

Vincke, P (1986) Analysis of multicriteria decision aid in Europe, *European Journal of Operational Research,* Vol. 25, pp. 160-8.

Von Winterfeldt, D and Edwards W (1986) *Decision Analysis and Behavioral Research*, Cambridge University Press, Cambridge.

Watson, S R and Buede D M (1987) *Decision Synthesis: The Principles and Practice of Decision Analysis*, Cambridge University Press, Cambridge.

Weber, M (1987) Decision making with incomplete information, *European Journal of Operational Research,* Vol. 28, pp. 44-57.

Weber, M, Eisenfuhr F and Von Winterfeldt D, (1987) *The Effects of Splitting Attributes on Weights in Multiattribute Utility Measurement*, Mimeo, RWTH, Aachen.

Wilcox, W B (1976) (ed.) *The Papers of Benjamin Franklin*, Volume 20, Yale University Press, New Haven.

Yoon, K (1980) *Systems Selection by Multiple Attribute Decision Making*, Ph.D. dissertation, Kansas State University, Manhattan, Kansas.

Zahedi, F (1986) The analytic hierarchy process - a survey of the method and its applications, *Interfaces*, Vol. 16, No. 4, pp. 96-108.

APPENDIX 1

COMPASS CODE	NAME	UNITS OF MEASUREMENT/SCALE	SUBJECTIVE (S) "OBJECTIVE" (O) SCALE	SCORE MINIMUM	SCORE MAXIMUM	NUMBERS AFFECTED MINIMUM	NUMBERS AFFECTED MAXIMUM	SCALING FACTOR	WEIGHT	ENTER
A	SAFETY									
A1	Personal injury accidents									
A1.1	Vehicle only									
A1.1.1	Slight)		O	0	10	-	-	0.1	0.004	#WO/#W
A1.1.2	Serious)		O	0	5	-	-	0.2	0.024	As A1.1.1
A1.1.3	Fatal)	In each of the six categories, number of accidents per year	O	0	2	-	-	0.5	0.162	As A1.1.1
A1.2	Pedestrian)									
A1.2.1	Slight)		O	0	10	-	-	0.1	0.004	As A1.1.1
A1.2.2	Serious)		O	0	5	-	-	0.2	0.024	As A1.1.1
A1.2.3	Fatal)		O	0	2	-	-	0.5	0.162	As A1.1.1
A2	Highway characteristics									
A2.1	Standard of carriageway structure	Score x flow (vehicles) p.a. 0 = very good 2 = good 4 = satisfactory 6 = inadequate 8 = poor 10 = very poor	S	0	10	0	10^{6}	10^{-7}	0.016	[Flow of vehicles p.a.& sco WO/W
A2.2	Degree to which existing highway characteristics fall below current standards	Score x flow (vehicles) p.a. 0 = not at all 2 = very little 4 = slightly 6 = moderately 8 = substantially 10 = very substantially	S	0	10	0	10^{5}	10^{-7}	0.020	As A2.1
A2.3	Standard of bridges, culverts and other structures	As A2.1	S	0	10	0	10^{6}	10^{-7}	0.016	As A2.1
A2.4	Standard of existing pedestrian facilities	As A2.1, but flow measured as number of pedestrians	S	0	10	0	10^{5}	10^{-7}	0.020	As A2.1
B	TRAFFIC									
B1	Travel time									
B1.1	Cars)		O	0	1	-	-	1	0.002	000's vehicle WO/W
B1.2	LCV's)		O	0	1	-	-	1	0.004	As B1.1
B1.3	HCV's)	Thousands of vehicle hours per year	O	0	1	-	-	1	0.005	As B1.1
B1.4	Public transport)		O	0	1	-	-	1	0.002	As B1.1
B1.5	Cyclists)		O	0	1	-	-	1	0.002	As B1.1
B1.6	Pedestrians)		O	0	1	-	-	1	0.002	As B1.1
B2	Operating costs									
B2.1	Cars)		O	0	1	-	-	1	0.002	£000's W
B2.2	LCV's)	Thousands of pounds per year	O	0	1	-	-	1	0.002	As B2.1
B2.3	HCV's)		O	0	1	-	-	1	0.002	As B2.1
B2.4	Public transport)		O	0	1	-	-	1	0.002	As B2.1
B3	Delay during construction)	Thousands of vehicle hours	O	0	1	-	-	1	0.002	000's vehicle W
C	ENVIRONMENT									
C1	Noise/vibration	365 x Σ [number of people suffering daily noise level (d_i) dBA>60dBA x (d_i - 60)]	O	60	80	0	10^{3}	0.00005	0.071	Total pers... dBA>60 W
C2	Pollution	365 x Σ [number of people suffering daily CO pollution level (p_i) ppm>4ppm x (p_i - 4)]	O	4	12	0	10^{3}	0.000125	0.040	# of peo affected CO level 4ppm WO/
C3	Severance/conflict									
C3.1	Agricultural land	Σ[score x area (ha) for each farm affected] 0 = none 2 = minor 4 = significant 6 = moderate 8 = severe 10 = very severe	S	0	10	0	500	0.0002	0.009	Total Hectare score WC
C3.2	Urban severance	Score x number of people affected (scores as for C3.1)	S	0	10	0	10^{4}	10^{-5}	0.048	[# of people affected score]
C3.3	Pedestrians' and cyclists' fear and danger	As C3.2	S	0	10	0	10^{4}	10^{-5}	0.028	As C3.2
C4	Visual disamenity									
C4.1	Intrusion (traffic related)	As C3.2	S	0	10	0	10^{3}	10^{-4}	0.029	As C3.2
C4.2	Obstruction (infrastructure)	As C3.2	S	0	10	0	10^{3}	10^{-4}	0.034	As C3.2
C5	Disruption to residents and the public during construction	As C3.2	S	0	10	0	10^{3}	10^{-4}	0.006	[# of people affected & score
D	PLANNING AND DEVELOPMENT									
D1	Housing	Annual person spaces Σ[number of houses built x average occupancy]	O	0	100	-	-	10^{-3}	0.057	# of person spaces
D2	Employment	Net change in annual jobs	O	0	100	-	-	10^{-2}	0.116	# of jo
D3	Planning/policy relevance	-2 or 2 = minor -4 or 4 = significant -6 or 6 = moderate -8 or 8 = major -10 or 10 = very important/severe (negative for violations; positive for furtherance)	S	0	(±)10	-	-	0.1	0.085	
E	SCHEME COSTS									
E1	Capital	Thousands of Pounds	O	0	-	-	-	1	1	
E2	Maintenance	As E1	O	0	-	-	-	12	1	
E3	Grant	As E1	O	0	-	-	-	1	1	

APPENDIX 2 - DERIVATION OF ATTRIBUTE WEIGHTS USING THE ANALYTIC HIERARCHY APPROACH

A questionnaire was prepared in which respondents were presented initially with seven clusters of lowest-level attributes. For each cluster the respondent has to estimate w^i/w_j ratios for all $i < j$. For example:

	A1.1.1	A1.1.2	A1.1.3	A1.2.1	A1.2.2	A1.2.3
A1.1.1	1	½	¼	1	½	¼
A1.1.2	-	1	½	2	1	½
A1.1.3	-	-	1	4	2	1
A1.2.1	-	-	-	1	½	¼
A1.2.2	-	-	-	-	1	½
A1.2.3	-	-	-	-	-	1

In each case, the weight comparison elicited relates to the value of a change from the minimum possible level and the lowest number of people affected to the maximum level and the highest number of people for attribute i to an equivalent change for attribute j.

This information for each of the clusters was then processed by the *Expert Choice* software which implements the AHP process and calculates estimates of the relevant w_j (normalised within each cluster) using Saaty's original eigenvector method. The weights derived from the matrix shown above, were:

	A1.1.1	A1.1.2	A1.1.3	A1.2.1	A1.2.2	A1.2.3
Weight	.071	.143	.286	.071	.143	.286

The full set of weights for all 32 attributes is computed first by repeating the above exercise for each of the lowest-level clusters and then by eliciting further comparisons hierarchically in which representatives from each of the clusters are successively compared with each other. In this case the process developed as follows:

	A2.1	A2.2	A2.3	A2.4	Derived Weight
A2.1	1	1	2	$\frac{1}{4}$	0.159
A2.2	-	1	2	$\frac{1}{4}$	0.159
A2.3	-	-	1	$^1/_6$	0.086
A2.4	-	-	-	1	0.595

	B1.1	B1.2	B1.3	B1.4	B1.5	B1.6	
B1.1	1	1	1	1	1	1	0.167
B1.2	-	1	1	1	1	1	0.167
B1.3	-	-	1	1	1	1	0.167
B1.4	-	-	-	1	1	1	0.167
B1.5	-	-	-	-	1	1	0.167
B1.6	-	-	-	-	-	1	0.167

	B2.1	B2.2	B2.3	B2.4	
B2.1	1	1	1	1	0.250
B2.2	-	1	1	1	0.250
B2.3	-	-	1	1	0.250
B2.4	-	-	-	1	0.250

	C3.1	C3.2	C3.3	
C3.1	1	$\frac{1}{2}$	$^1/_5$	0.122
C3.2	-	1	$^1/_3$	0.230
C3.3	-	-	1	0.648

	C4.1	C4.2	
C4.1	1	1	0.500
C4.2	-	1	0.500

	D1	D2	D3	Derived Weight
D1	1	½	4	0.323
D2	-	1	6	0.588
D3	-	-	1	0.089

	A1.1.1	A2.1	
A1.1.1	1	6	0.857
A2.1	-	1	0.143

	B1.1	B2.1	B3	
B1.1	1	2	6	0.588
B2.1	-	1	4	0.323
B3	-	-	1	0.089

	C1	C2	C3.1	C4.1	C5	
C1	1	4	4	6	7	0.515
C2	-	1	3	3	6	0.236
C3.1	-	-	1	1	5	0.122
C4.1	-	-	-	1	1	0.077
C5	-	-	-	-	1	0.049

	A1.1.1	B1.1	C1	D1	
A1.1.1	1	3	3	4	0.507
B1.1	-	1	1	3	0.204
C1	-	-	1	3	0.204
D1	-	-	-	1	0.085

The resulting set of weights, normalised to one is:

Attribute	Weight	Attribute	Weight
A1.1.1	.049	A2.1	.008
A1.1.2	.098	A2.2	.008
A1.1.3	.196	A2.3	.004
A1.2.1	.049	A2.4	.030
A1.2.2	.098		
A1.2.3	.196		
B1.1	.020	B2.1	.011
B1.2	.020	B2.2	.011
B1.3	.020	B2.3	.011
B1.4	.020	B2.4	.011
B1.5	.020		
B1.6	.020	B3	.003
C1	.020	C3.3	.025
C2	.009	C4.1	.003
C3.1	.005	C4.2	.003
C3.2	.009	C5	.002
D1	.008		
D2	.015		
D3	.002		

Using the geometric mean to calculate the weights in each cluster requires that the complete matrix be employed, where $a_{ij} = (a_{ji})^{-1}$. For example, completing the first cluster's matrix yields the following:

						$\left[\prod_j a_{ij}\right]^{1/6}$	w_i
1	$\frac{1}{2}$	$\frac{1}{4}$	1	$\frac{1}{2}$	$\frac{1}{4}$	0.500	0.071
2	1	$\frac{1}{2}$	2	1	$\frac{1}{2}$	1.000	0.143
4	2	1	4	2	1	2.000	0.286
1	$\frac{1}{2}$	$\frac{1}{4}$	1	$\frac{1}{2}$	$\frac{1}{4}$	0.500	0.071
2	1	$\frac{1}{2}$	2	1	$\frac{1}{2}$	1.000	0.143
4	2	1	4	2	1	2.000	0.286

In this case, the geometric mean procedure yields identical weights to those derived from the eigenvector. This occurs because (unusually) the ratios in this matrix are entirely consistent with each other throughout (Crawford, 1987). More generally, there will be some differences, but often small ones. For example, the weights derived for the penultimate of the matrices analysed earlier are:

	C1	C2	C3.1	C4.1	C5
Weight	0.518	0.234	0.118	0.079	0.048

Relative to the potential inaccuracies inherent in the rest of the PAT process, the differences are trivial, although it should be borne in mind that the hierarchical derivation of the full weight set in COMPASS will cause some errors to be magnified as different weight estimates are successively combined.

APPENDIX 3 - COMPUTATION OF DUMMY SCORES FOR CIRCUMSTANCES IN WHICH ONE LOWEST-LEVEL ATTRIBUTE IS CHOSEN TO REPRESENT ITS CLUSTER OF ATTRIBUTES

Suppose that attribute 1 has been chosen to represent all the m attributes in its cluster. What is required is to input dummy scores for the remaining (unevaluated) attributes in its cluster which, after scaling but before weighting, yield the same scaled score as that recorded by the chosen attribute.

Let ΔS_i $(i = 1,...,m)$ be the net change in score on attribute i and let SF_i be the corresponding scaling factor (see appendix 1). Then, we require that

$$\Delta S_1 \times SF_1 = \Delta S_i \times SF_i$$

$$\text{i.e.} \qquad \Delta S_i = \Delta S_1 \times \frac{SF_1}{SF_i} \quad (i = 2,...m)$$

Thus, to obtain the required computational result from COMPASS, all that is necessary is as follows:

(a) For attributes that are not multiplied by a flow measure:

- (i) set the score without the scheme in place to zero;
- (ii) set the score with the scheme in place to

$$\Delta S_i = \frac{SF_1}{SF_i} \times \Delta S_1$$

(b) For attributes that are multiplied by a flow factor:
Let N_i^w and N_i^{wo} be respectively the flows with and without the scheme; let S_i^w and S_i^{wo} be project scores with and without the scheme.

For projects where a higher score connotes an improvement, set S_i^{wo} and N_i^{wo} to zero. Then enter S_i^{w} and N_i^{w} such that

$$N_i^{w} S_i^{w} = \frac{SF_i}{SF_i} (N_1^{w} S_1^{w} - N_1^{wo} S_1^{wo})$$

For projects where a higher score connotes a deterioration, the roles of (N_i^{w}, S_i^{w}) and (N_i^{wo}, S_i^{wo}) should be reversed.

APPENDIX 4 - COMPUTATION OF DUMMY SCORES FOR CIRCUMSTANCES IN WHICH A CLUSTER OF ATTRIBUTES IS SCORES SUBJECTIVELY

It is necessary to ensure thall all lowest-level attributes in the cluster in question achieve the same scaled score as has been estimated for the cluster as a whole.

Suppose the group of attributes is scores at S ($0 \leqslant S \leqslant 10$). For each lowest-level attribute(i) within its cluster, using appendix 1:

(a) Compute the range [(Maximum Score x Maximum Number affected) - (Minimum Score x Minimum Number affected)];

(b) Estimate F = 10S% of this range;

(c) For projects where a high score connotes improvement, set S_i^{wo} and N_i^{wo} to zero (see appendix 3 for notation). Then enter S_i^{w} and N_i^{w} such that $N_i^{w} S_i^{w} = F$;

(d) For projects where a low score connotes improvement, reverse the roles of (S_i^{wo}, N_i^{wo}) and (S_i^{w}, N_i^{w}).