Australian Government
Department of Defence
Defence Science and
Technology Organisation

Uncertainties in the Analytic Hierarchy Process

Lewis Warren

Command and Control Division
Information Sciences Laboratory

DSTO-TN-0597

ABSTRACT

When choosing a decision analysis technique to model a particular complex decision the fundamentals of the technique chosen should be understood by the analyst, and they should be appropriate for the characteristics of the problem itself. For analysing such complex decisions the Analytic Hierarchy Process is one of the most commonly used techniques. This technical note highlights a number of theoretical issues, some well-known and others less well-known, that introduce a considerable degree of uncertainty into the computed output priorities for the decision alternatives.

Published by

DSTO Information Sciences Laboratory
PO Box 1500
Edinburgh South Australia 5111 Australia

Telephone: (08) 8259 5555
Fax: (08) 8259 6567

AR-013-275
December 2004

Uncertainties in the Analytic Hierarchy Process

Executive Summary

Among the bewildering array of decision analysis techniques that apply systematic and structured analysis to complex decisions, the Analytic Hierarchy Process (AHP) is one of the most widely used. From the early days it has been noted that it can result in certain anomalies, the rank reversal problem being the most widely known. Whether or not these behavioural anomalies are actually reflected in real-world decision makers has been a topic of hot discussion. In the case of rank reversal, many authors believe that it is valid in certain real-world situations; but besides rank reversal there are other anomalies that are harder to justify. When choosing a decision analysis technique to model a particular complex decision, the fundamentals of the technique should be understood by the analyst, and they should be appropriate for the characteristics of the problem itself. When a problem cannot be decomposed into independent facets, for example, then a model that requires criteria independence such as the AHP should not be applied.

However, there may still be a number of candidate techniques to choose from and it would be prudent to choose one that is robust, not overly simplistic, and is based on sound computational methods. This technical note has been motivated by the widespread usage of the AHP without generally acknowledging, or appreciating, the uncertainties embedded in its results. Frequently the justification for adopting the AHP seems to be the belief that any systematic method will do, because in the end, the primary purpose is to help the decision maker establish the elemental structure of the problem, without the necessity to assume that the numerical outputs are exact. Furthermore, it is also relatively easy to explain the AHP hierarchical decomposition model to most non-technical customers. While hierarchical decomposition and aggregation is a natural approach to many problems, the fact is that there are many questionable theoretical issues in the AHP technique. Over the last twenty years several authors have commented on these difficulties and many of those criticisms have been summarised in this technical note. In addition, some less well-known issues are also highlighted.

An overview of the foundations of the AHP is initially provided and the successive assumptions upon which the computational methods are based are discussed. The conclusion of this investigation is that there is a combination of questionable procedures in the technique, such that there is always a significant degree of uncertainty surrounding the output priorities of the method. A decision maker needs to be aware of such issues if an appropriate method is to be selected and correctly applied.

Contents

1. Introduction

The Analytic Hierarchy Process (AHP) was developed by Saaty around 1970 and the first application of it to a real-world problem was in 1973. The method was published [24] in 1980 and since then it has become one of the most widely applied techniques for decision analysis. Among the many reasons for this are the existence of user-friendly software with built-in sensitivity analysis, the apparent mathematical sophistication of the technique, and the immediate attraction of hierarchically structured decisions. Apart from the mathematical details, the overall concept is also relatively easy to explain to non-technical managers. Two important features of the method are the elicitation of subjective ratings for pairwise comparisons of factors, and the hierarchical aggregation of priorities derived from the pairwise comparison matrices of ratings, into a global vector of priorities for decision alternatives. Over the years a number of authors have investigated the computational methods of the AHP and raised concerns about the validity of some of the assumptions upon which they are based.

This technical note summarises these concerns, highlighting the relevant assumptions that are usually given only tacit acknowledgment. The range of concerns is divided into two categories for simplicity: primary problems which affect the axiomatic foundations of the method, and secondary problems which are either behavioural symptoms of deeper problems, or else context dependent and not always of significant concern. The well-known rank reversal problem, whereby the addition of a new alternative changes the priorities of the other alternatives, is included in the secondary problem category. A major difficulty when evaluating the relative merits of a structured decision theoretic is that the outputs cannot in general be validated in any absolute way. In a sense they are all speculative. While practical aspects can be compared, and their consistency compared through simulation over diverse input sets, the validity of the computed measures cannot be absolutely ascertained; all this assuming that the techniques being compared are suitable for the target problem being used. So the selection of a method should be based on a reasonable understanding of the computational details and their respective assumptions and limitations. It is also generally desirable that a method should be robust, meaning that the underlying assumptions are reasonable, as well as the outputs not being overly sensitive to small changes in inputs. In addition to these considerations it also helps if the method is not too complex so that it can be explained to the non-technical decision makers.

The objective of this technical note is to highlight some features of the AHP computations that are based on questionable assumptions. Individually, these features can introduce significant amounts of uncertainty into the computed priority measures for decision alternatives. However in combination, uncertainty is magnified to the extent that there is considerable doubt surrounding the computed priority measures which limits their usefulness in making decisions. The work in this report was conducted under the Strategic Operations Division sponsored task JTW 02/304 "Information Operations Experimentation", and is important for the development of an operations evaluation framework.

2. An Overview of the AHP

2.1 Saaty's Axioms

The following extracts from Saaty [28, pp. 841-842] provide an overview of the AHP:

"The AHP is a systematic procedure for representing the elements of any problem. It organizes the basic rationality by breaking down a problem into smaller constituent parts and then calls for only simple pairwise comparison judgments to develop priorities in each hierarchy. There are three principles which one can recognize in problem solving. They are the principles of decomposition, comparative judgments, and synthesis of priorities."

"The decomposition principle calls for structuring the hierarchy to capture the basic elements of the problem … from the more general (and sometimes uncertain) to the more particular and definite. One can then start at the bottom, identifying alternatives for that level and attributes under which they should be compared which fall into the next level up ... In general. the bottom level of the hierarchy contains the resources to be allocated, or the alternatives from which the choice is to be made."

"The principle of comparative judgments calls for setting up a matrix to carry out pairwise comparisons of the relative importance of elements in the second level with respect to the overall objective (or focus) of the first level… Additional comparison matrices are used to compare the elements of the third level with respect to the appropriate parents in the second, and so on down the hierarchy... The next step deals with the composition of the derived ratio scales."

"Priorities are synthesized from the second level down by multiplying local priorities by the priority of their corresponding criterion in the level above, and adding them for each element in a level according to the criteria it affects... This gives the composite or global priority of that element which is then used to weight the local priorities of elements in the level below compared by it as criterion (sic), *and so on to the bottom level."*

As the lowest strata of assumptions, axioms provide the foundations for any methodology or technique. Saaty [28] has specified four axioms for the AHP somewhat ambiguously, and these have been described more simply by Forman and Gass [15] as follows.

- First, the *reciprocal* axiom requires that if $P_C(A,B)$ is a paired comparison of elements A and B with respect to their parent element C, representing how many times more element A possesses a property than does element B, then $P_C(B,A) = 1/ P_C(A,B)$.

- Second, the *homogeneity* axiom states that the elements being compared should not differ by too much in the property being compared. To prevent large errors in judgments the measures corresponding to the linguistic ratings should be limited to an order of magnitude.

- Third, the *synthesis* axiom states that judgment about the priorities of elements in a hierarchy should not depend on lower level elements. This axiom is required for hierarchic composition to apply and apparently means that the importance of higher level objectives should not depend on the priorities or weights of any lower level factors. (This is slightly different to stating that all factors should be independent for additive priority aggregation.)

- A fourth *expectation* axiom says that individuals who have reasons for their beliefs should make sure that their ideas are adequately represented for the outcomes to match these expectations. This axiom means that output priorities should not be radically different to any prior knowledge or expectation that a decision maker has.

2.2 The Analytical Process

The AHP is fundamentally an additive weighted aggregation of priority scores that have been derived from subjective scores for pairwise comparisons of lowest level factors or criteria. An example of a military application [18] is the determination of prioritised alternative solutions for an upgrade to an Airborne Surveillance System.

2.2.1 Hierarchical Decomposition of the Decision

The decision is first decomposed into a hierarchical structure of the necessary and sufficient set of elements or factors that are needed when making the decision. Figure 1 shows such a hierarchical structure of factors where A is the overall priority of an alternative based on all factors needing consideration, and higher level factors B,C,D can be cognitive categories of factors such as Lifecycle Cost, Benefits, and Risks. The lower level factors (as ellipses in Figure 1) then represent the sub-factors or sub-criteria within each factor category.

To explain the method and the computational stages, we will loosely use the military AHP application referred to above. The following five alternatives are candidates selected for consideration to upgrade the airborne surveillance capability.

Alt1: Airborne Warning & Control System, Alt2 : Strategic Surveillance System
Alt3: Tactical Surveillance System Alt4: Electromagnetic Intelligence
Alt5: Battlefield Surveillance System

The top level categories are:
B = Acquisition Feasibility, C = Sustainability, D = Military Gain

The next level factors are:

B1 = Availability, B2 = Cost

C1 = Import Content, C2 = Technology Absorption, C3 = Indigenous Production

D1 = Non-existing Capability, D2 = Enhancement of Exis ting Capability,
D3 = Reduction in Enemy Capability, D4 = Morale Booster for Own Force.

And the lowest level factors are:

B21 = Set-up Cost, B22 = Running Cost, B23 = Annual Equivalent Cost

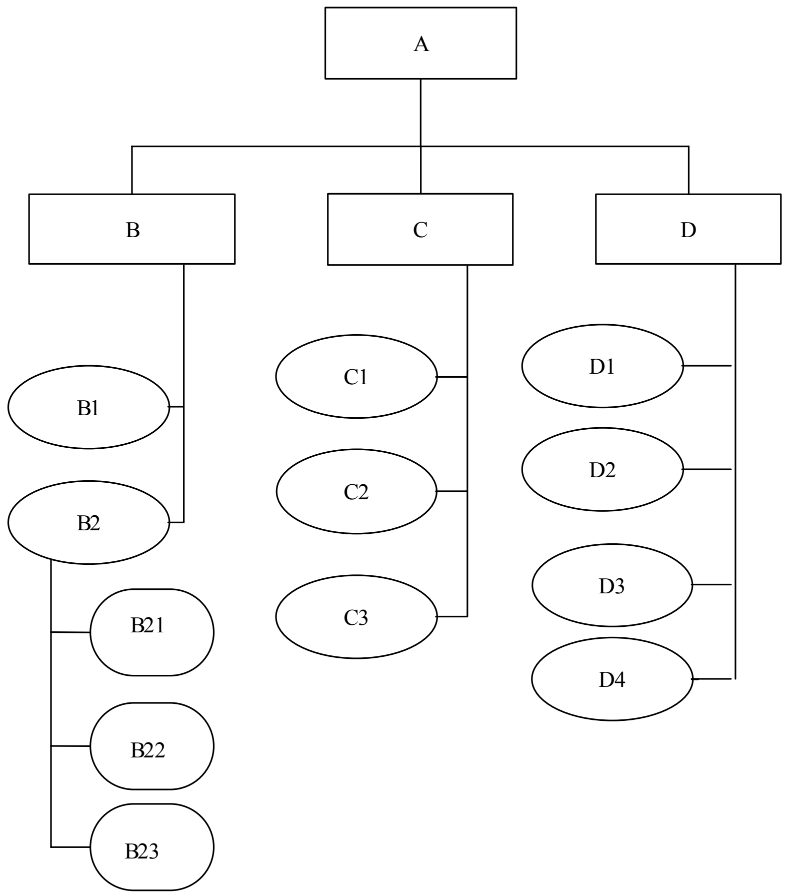

Figure 1: An Example of a Decision Factor Hierarchy

2.2.2 Pairwise Comparisons

Two types of pairwise comparisons are made in the AHP. The first is between factor pairs within the same hierarchical level and involves analyst input of relative importance ratings. The computed measures from these inputs are called factor weights and used in the final hierarchical merit aggregation process. Factor weights are determined from top-down factor comparisons and scaled so that the sum of weights under any node equals one. The second type of pairwise comparison is between pairs of alternatives and is used to determine their relative merits against each leaf or terminal node.

To make all such pairwise comparisons, input ratings of relative strength require some sort of graded rating scale.

2.2.2.1 Selection of Comparison Scale Type

Since the meaning of a numerical rating is determined by the type of scale it is based upon (e.g. ordinal, interval or ratio), the scale type must be established. Saaty [27] states that the AHP uses a ratio scale:

" relative measurement is a method for deriving ratio scales from paired-comparisons representing absolute numbers."

This will subsequently be discussed in more detail.

2.2.2.2 Selection of Comparison Scale Units

When using a ratio scale for mutual comparisons, the numbers represent the relative magnitude of the property possessed by the two factors being compared. However, in the AHP decision maker inputs are usually in the form of verbal or linguistic ratings of relative importance, and these are then converted to numbers in the comparison matrix. For example, the scale commonly used as proposed by Saaty is as follows.

Equal Importance,	Mildly Stronger,	Stronger,	Much Stronger	Extremely Stronger
(1),	(3),	(5),	(7),	(9)

Thus if a decision maker answers that the importance of D (Military Gain) is "Stronger" than C (Sustainability) it would be converted to a numerical value of 5 in the comparison matrix. Intermediate numbers then correspond to intermediate grades and although a variety of other numerical conversion scales have also been proposed, this is the original or standard AHP scale.

The reciprocal of the numbers then corresponds to the factor comparison inverted (as in the *reciprocity axiom*) and the set of measures corresponding to the above 5 grades is:

$$\{ 0.11,\ 0.14,\ 0.20,\ 0.33,\ 1,\ 3,\ 5,\ 7,\ 9 \}$$

Example 1: The relative importance of factors D, B, C with respect to A.

	Military Gain (D)	Acquisition (B)	Sustainability (C)
Military Gain	**1**	**3**	**5**
Acquisition	**1/3**	**1**	**2**
Sustainability	**1/5**	**1/2**	**1**

Example 2: The relative importance of sub-factors D1, D2, D3, D4 with respect to D.

	D1	D2	D3	D4
D1	**1**	**1/5**	**1/4**	**1/4**
D2	**5**	**1**	**2**	**3**
D3	**4**	**1/2**	**1**	**3**
D4	**4**	**1/3**	**1/3**	**1**

Example 3: The relative merit of Alternatives with respect to AVAILABILITY (B1)

	Alt1	Alt2	Alt3	Alt4	Alt5
Alt1	**1**	**1/7**	**1/3**	**1/5**	**1/9**
Alt2	**7**	**1**	**2**	**1**	**1/2**
Alt3	**3**	**1/2**	**1**	**1/2**	**1/3**
Alt4	**5**	**1**	**2**	**1**	**1/2**
Alt5	**9**	**2**	**3**	**2**	**1**

In Example 3 the merit or performance of Alternative 2 is Much Stronger (7) than that of Alternative 1 with respect to AVAILABILITY.

2.2.3 Pairwise Matrix Evaluation

From the square matrices of pairwise comparison ratings, the AHP determines the factor weights and alternative priorities using a method based on matrix algebra eigenvalue techniques.

For any square matrix A, λ is an eigenvalue associated with a vector Ψ such that $A\ \Psi = \lambda\ \Psi$, where Ψ is called the corresponding eigenvector.

Step 1: Find the largest eigenvalue λ that solves the characteristic polynomial for the above equation.

Step 2: Calculate the corresponding principal eigenvector for the maximum eigenvalue.

Step 3: Normalise the principal eigenvector values so that the elements sum to unity. This normalisation is by the "city-block" method, where the normalisation constant is the sum of the elements. Then the normalised elements represent either the relative weights of factors, or the relative priorities of alternatives against a criterion.

Using this procedure the computed factor weights (for the Example 1 and 2 matrices) and alternative priorities (for the Example 3 matrix) are as follows.

Relative Weights of Factors:

D: Military Gain (0.649), B: Acquisition Feasibility (0.229), C: Sustainability (0.122).

D1: Non-existing Capability (0.066), D2: Enhance Own capability (0.458),
D3: Reduction in Enemy Capability (0.312), D4: Morale Boosting (0.164)

Relative Priorities of Alternative with respect to Availability:

Alt1 (0.039), Alt2 (0.230), Alt3 (0.118), Alt4 (0.215), Alt5 (0.398)

2.2.4 Additive Weighted Aggregation of Priorities

When all relative weights of factors have been determined, and relative priorities for alternatives determined for each of the terminal factors, weighted priority aggregation occurs through the hierarchy from the bottom to the top.

The aggregate priority at each node for an alternative is the additive weighted sum of its children's priorities ($\sum_{1}^{n} w(i)\ prioritity(i)$). The aggregate global priority vector at the top (A) then represents relative preference measures for alternatives over all factors in the decision, and the ranking of these measures determines the relative preference order of the alternatives.

2.2.5 Evaluation of Rating Inconsistency

One appealing feature of the AHP is the ability to evaluate pairwise rating inconsistency. The eigenvalue technique enables the computation of a consistency measure which is an approximate mathematical indicator of the inconsistencies or intransitivity in a set of pairwise ratings. This consistency measure is a function (called the Consistency Index) of the maximum eigenvalue and the size of the square matrix. Then, if the ratio of the Consistency Index to a similar index derived by assuming that the pairwise comparisons had been generated by a random process, is greater than 0.1 (or 10%) the level of inconsistency in the set of ratings is considered to be unacceptable. In this situation a review or repeat of the ratings is then recommended.

3. Some Problematic Features

3.1 Primary Problems

3.1.1 Criticism of Saaty's Axioms

Some authors have questioned the adequacy of Saaty's axioms. Barzilai [6] states:

"…the axioms underlying the AHP are meaningless as well. If they do not properly characterise the AHP, they are of no interest. On the other hand, if they do, they cannot be meaningful either, since they characterize a methodology which suffers from multiple flaws."

Axiom 1: The reciprocal axiom.
This states a necessary mathematical requirement, essential for ratio scale measures of the ratios of the importance property of two factors. For true ratio scale measures the axiom is valid, but some authors suggest that subjective ratings of relative importance cannot be measured on a scale with an absolute zero since subjective importance cannot be quantified exactly. So the validity of this axiom actually depends on whether or not a ratio scale is actually applied, and this is questionable as will be subsequently discussed.

Axiom 2: The order of magnitude rating limit
Saaty states that when the difference in importance of two factors is very great meaningful comparisons cannot be made. For this reason, a limit of one order of magnitude is applied, or 10 on a decimal scale. Thus, Saaty uses 9 on his recommended scale as the maximum rating. When the difference in property magnitudes is significantly greater than this Saaty recommends the definition of different elements and clusters of elements i.e. to readjust the hierarchical decomposition. However, it may not be practicable or desirable to redefine a model

when such a divergence in priority values is encountered. In these cases, the upper limit is effectively 9 and greater ratio values are lost. (Similarly for the lower limit.)

Axiom 3: Ratings on any level are not affected by lower level ratings.
This type of independency is additional to the factor or criteria independency that is required for additive hierarchical aggregation of priorities.

Axiom 4: The expectation axiom
This states that results must comply with the decision maker's belief or intuition. It is defined to exclude any results that may appear irrational and be caused by a crude, incomplete, or false model.

Overall, this is an adhoc set of axioms. Axioms 1 and 3 do address mathematical foundations to some extent, but axiom 2 is simply a constraint based on quasi-empirical evidence, while axiom 4 is a posthoc condition and does not address mathematical foundations at all. So there is some concern that this set of axioms does not comprise a necessary and sufficient set of mathematical prerequisites as the foundations of a computational methodology should. This concern is reinforced by ratings not being true ratio scale measures (to be discussed), which negates the validity of axiom 1.

3.1.2 Misunderstanding of the Rating Scale Type

The foundations of Measurement theory have been described by Stevens [38], Roberts [22], and others [7] [19] [37]. Simply speaking there are three types of scales which numerical measures may be based upon: ordinal, interval and ratio scales. The amount of information embedded by the scale type increases from the minimum in ordinal measures to the maximum in ratio scale measures. Admissible algebraic operations on measures must accord with the amount of information embedded in a scale. Ratio scales must possess an absolute zero which then enables division and multiplication, as well as addition and subtraction. Division and multiplication of individual measures is not permitted for interval scale measures because they have no absolute zero. However, subtraction, averaging, and ratios of differences of interval scale measures for the same concept are permitted. No algebraic operations are admissible on ordinal scale measures. Stevens [38] has pointed out that measurement fundamentals are often neglected by scientists because they frequently work in the physical domain where the ratio scale is implicit. However, for psychological measurement and subjective evaluations of qualitative variables without measurable properties, as used in decision analysis, careful attention should be given to implicit scale type since it can limit the range of justifiable and realistic mathematical operations.

In the AHP literature there is considerable ambiguity as to whether the input relative importance ratings are on an *implicit* ratio scale, or whether the derived priorities computed from the comparison matrix are on a *derived* ratio scale. Both understandings seem to exist.

Saaty in 1993 [29, p. 1]
"Relative measurement is a method for deriving ratio scales from paired comparisons represented by absolute numbers."

Saaty in 1996 [30, p. 34]
"By now the reader has seen how we derive a ratio scale from numerical dominance (as distinct from profile, proximity, or conjoint) paired comparison matrix. Actually it does not matter what numbers we use, we always get a ratio scale as the principal eigenvector..."

After this Saaty defines four kinds of ratio scales: absolute ratio scale, ratio ratio scale, ordinal ratio scale, and chaotic ratio scale and these definitions highlight Saaty's basic misunderstanding of the ratio scale concept.

Saaty in 2001 [31, p. 4]
"In using the AHP to model a problem, one needs a hierarchic or network structure to represent that problem, as well as pairwise comparisons to establish relations within the structure. In the discrete case these comparisons lead to dominance matrices and in the continuous case to kernels of Fredholm Operators from which ratio scales are derived in the form of principal eigenvectors, or eigenfunctions, as the case may be."

If there was ratio scale ambiguity in Saaty's initial explication of the AHP in 1980, it would seem clear that currently he adopts the "derived" ratio scale interpretation, and this is also reflected in the recent literature of other authors of the AHP school.

For example, Forman and Gass [15, p. 470] in their 2001 "State-of-the-AHP–art" exposition:

"Ratio measure is necessary to represent proportion and is fundamental to physical measurement. This recognition, plus a need to have a mathematically correct, axiomatic-based methodology, caused Saaty to use pairwise comparisons of the hierarchical factors to derive (rather than assign) ratio-scale measures that can be interpreted as final ranking priorities (weights). "

Saaty also states that the AHP produces ratio scale measures for priorities regardless of whether input information is objective or subjective information. However, earlier literature from other members of the AHP school illustrates a different understanding.

For example, Harker and Vargas [17, p. 1389] in 1987 :
"Saaty (1980) has proposed that we use a ratio scale between 1 and 9, although as we have discussed, this scale is open to debate."

(Note that in this quote the debate being referred to is about the units or grades of the scale and not about the scale type.)

With this understanding, when A is Mildly Stronger than B, and is assigned 3 on Saaty's scale, the meaning for a ratio scale measure is :

A Importance = 3 x B Importance.

In fact, the input ratings for comparisons must be interpreted in this way if the subsequent mathematical operations are to be admissible.

It is also apparent that some members of the AHP school assume that since the input ratings {1,3,5,7,9} are of the strength of a quotient (A/B), then *ipso facto* they automatically induce ratio scale priority measures.

For example, Forman and Gass [15, p.482] in their 2001 exposition of AHP:
" Because each pairwise comparison is already a ratio, the resulting priorities will be ratio-scale measures as well."

This is false because the type of scale induced is determined by the nature of the variable being considered and not whether it *represents* a ratio or not. For example, "Height of tree in metres" induces a ratio scale because the variable has an absolute zero. Different types of uncertainty are then embedded in this measure depending on how it is measured. Furthermore, the relative heights of two trees are also on a ratio scale.

Thus the variable being rated in the AHP is the quotient Importance A / Importance B and Saaty states [29] [30] that when ratios are rated the units of the variables cancel out so that a measure of the ratio of two non-measurable inputs, such as "importance", is automatically a ratio-scale measure.

This logic is also flawed because it is not the units of a scale that determine the presence of a ratio scale, but whether or not the property of the variable being measured has an absolute zero. Barzilai [6][7][8] has emphasised this point and has shown that both the existence and the location of an absolute zero comprise the necessary and sufficient conditions for the presence of a ratio scale.

There are two basic reasons why relative importance ratings can not have an absolute zero:

1. The "importance" of a decision element itself is context dependent and depends on the value system of an individual rater. In comparison, if the ratio of the height of two trees is also subjectively rated, it too would depend on the individual rater's characteristics (such as his eyes); but moreover, an absolute zero of the ratio is implied since it exists for the scale of "height" in a non-context dependent manner. Thus, subjective ratings of height ratios are also on a ratio scale, while they are not for importance ratios, especially when qualitative variables are involved.

2. Saaty's Axiom 2 states that the variables compared should be of the same order and consequently 9 is the maximum measure permitted in the AHP. He also states that a variable cannot be infinitely more (or less) important so that a relative importance measure of zero is non-existent. Thus, 0.11 (1/9) is the minimum measure possible in the AHP and an absolute zero does not exist for axiomatic reasons!

Various authors have questioned the validity of the assumption that the input ratings are ratio scale measures.

Stewart [39, p. 574] in 1992 :
" … the usual form of input required by the AHP is not the numerical ratio described above but rather a preference statement on a nominal nine point scale, which is interpreted as a ratio. … Justification for this quantitative interpretation of a nominal scale is anecdotal and has been questioned."

Barzilai [8, p.4] in 2001 :
"Since an absolute zero has not been established (and, in all likelihood, does not exist) for preference measurement, preference cannot be measured on ratio scales."

And also by AHP school members Forman and Gass [15]:
"The fundamental verbal scale is ordinal only because the intervals between the words on the scale are not necessarily equal. Despite the fact that the fundamental verbal scale used to elicit judgment is an ordinal measure, Saaty's empirical research showed that the principal eigenvector of a pairwise verbal judgment matrix often does produce priorities that approximate the true priorities from ratio scales such as distance, area and brightness. This happens because, as Saaty (1980) has shown mathematically, the eigenvector calculation has an averaging effect – it corresponds to finding the dominance of each alternative along all walks of length k, as k goes to infinity. Therefore, if there is enough variety and redundancy, errors in judgments, such as those introduced by using an ordinal verbal scale, can be reduced greatly."

Forman and Gass of the AHP school thus suggest that output priorities will be ratio scale measures for two reasons:
1. Because ratings are of ratios of something. (As discussed previously.)
2. Because they have been transformed from the input ordinal measures by the eigenvector calculation. (As in above quotation.)

There would seem to be little doubt that the input ratings, both verbal measures and their nominal numerical equivalents, are ordinal measures. No computational procedure on ordinal measures can add extra information to transform the input ordinal measures into output ratio scale measures. This is simply illogical. The inescapable conclusion is that the AHP performs inadmissible operations on ordinal measures, and therefore, the results of these computations, whatever they may be, are all meaningless.

3.1.3 A Weak Eigenvector Justification

This section is based on the assumption that the input ratings in the comparison matrix are on a ratio scale so that all algebraic operations would become admissible.

Many authors have questioned the justification for using the right hand principal eigenvalue and corresponding eigenvector. Saaty [23][26][27] argues that the "dominant" right eigenvector corresponding to the maximum eigenvalue should be used to estimate priorities because it can be used to estimate rating consistency.

Crawford and Williams [12, p. 388] comment on Saaty's reasoning thus:

" The argument is: The dominant eigenvector is a continuous function of the elements of the matrix, and, if the matrix is consistent, the eigenvector gives the unique (to within scalar multiplication) scale. Thus, if the elements of the matrix get perturbed slightly in the process of being subjectively quantified by a judge, the dominant eigenvector will return a scale only slightly different from the scale of an underlying consistent judgment matrix.

Although the classical analyst may worry about uniform continuity or other erudite intricacies of this argument, we are worried about a more basic oversight: the eigenvector is not the only continuous vector-valued function of judgment matrices that yields the correct scale when the matrix happens to be consistent. There are many others, including the vector of row sums, the vector of inverse column sums, any column of the matrix, and the whole ring generated by positive linear combinations of these and other solutions.

We are aware of the desirable properties of the eigenvector in characterizing a linear operator and its spectral decomposition, but the immediate relevance of these properties to this estimation problem seems open to question. "

Crawford and Williams also point out the benefits of using the Geometric Mean instead. Barzilai [3] as well highlights the benefits of the Geometric Mean and states that it is the *only* method for deriving weights from multiplicative matrices, as in the AHP, that satisfies fundamental consistency requirements.

Barzilai [2][3] suggests that the claim by Saaty and Vargas [25] that the right eigenvector "preserves rank strongly", implies that the left one does not. He demonstrates that both have the same rank preservation properties and that they can yield different rankings. Furthermore, he points out that there are infinitely many solutions that also have the same rank preservation properties. Barzilai [3] shows that the eigenvector solution depends on the description of the problem and the *arbitrary order of factor arrangement,* and he concludes that the justification for the eigenvector method is questionable.

While there have been several attempts using simulation [16][20][40] to assess the relative merits of different methods for comparison matrix evaluation, the results overall have not shown any method outperforming the others. Nevertheless, the fact that the Geometric Mean method (also called the logarithmic least squares method) can be applied with *incomplete* matrices is a practical advantage of that method.

In summary, no valid justification for using the eigenvector method can be found and more rigorous mathematical analysis suggests that the Geometric Mean would be preferable (if the inputs were in fact ratio scale measures).

3.1.4 Rating of "Relative Importance" of Criteria

In 1989 Schoner and Wedley [35] discussed the ambiguity that has been associated with criteria weights and showed how:

"there is a necessary correspondence between the manner in which criteria importances are interpreted and computed, and the manner in which the weights of the options under each criterion are normalized. In general, if this relationship is ignored, incorrect weights are generated for options under consideration regardless of whether new options are added or deleted. A rank reversal on the addition of an option is merely symptomatic of this fact."

And also in [35]:

"The problem arises in the generation of composite measures where there is measurement on more than one criterion. Paired comparisons at levels involving criteria must make reference to the magnitude of items in the immediate lower level but there is no requirement within conventional AHP for them to do so."

Thus, these authors suggest that criterion importance is not independent of alternative performance ratings and their normalisations of lower level criteria.

Dyer [13] has also identified this kind of dependency stating that weights of criteria are not independent of the performance measures on them, and if rated as if they were, the results are arbitrary. Although Dyer attempts to fix the problem by changing the normalisation procedure, Barzilai [3] has suggested that applying multi-level normalisation to priority or weight vectors is itself the problem.

Barzilai [5] reasons from the additive weighted aggregation of preference that:

$f(x) = \sum_{j=1}^{n} w_j x_j$ *implies* $\frac{w_i}{w_j} = \frac{\partial f}{\partial x_i} / \frac{\partial f}{\partial x_j}$, and this means that the weights must be dependent on the units of "x". In addition Barzilai also states that criteria weights should only apply to the terminal leaf-nodes of the criteria tree. This is not really a limitation since once the leaf-node weights have been determined, weights at upper level criteria are simply additively derived.

From these considerations, the authors above suggest that any top-down pairwise comparison of criteria relative importance will only yield arbitrary weight values.

3.1.5 Normalisation Anomalies

Normalisation is often applied to reduce measures of incommensurate variables to a dimensionless measure on the unit interval [0,1]. The common understanding is that these normalised measures can then be legitimately combined in algebraic operations.

As Forman states [14]:
"We cannot add numbers from different ratio scales and get meaningful results, but we can if the numbers belong to the same ratio scale. Normalization puts the priorities of alternatives appearing under different (sub)criteria on the same ratio scale, so that when we multiply by the weight of the corresponding (sub)criteria and add over all (sub)criteria, the result also belongs to the same ratio scale."

Not all types of normal measures of a set of numbers sum to unity, only those that have been produced with additive normalising constants. Multiplicative normalising constants produce normalised measures whose product is unity. Saaty uses the city-block method as an additive constant to derive normalised weights or priorities from comparison matrices that sum to unity.

Barzilai [2][4] [5] [6] suggests that normalisation is equivalent to rescaling and he has demonstrated that there can be problems underlying the mathematics of hierarchical aggregation of normalised measures. This hierarchical aggregation process, which Barzilai states was initially formulated by Miller [21] in 1966, is based on the concept of decomposition of criteria into a sub-criteria tree. In the AHP variant of Miller's process, Saaty unified Miller's multiple verbal scales into a single verbal scale.

Some of Barzilai's criticisms of Miller's hierarchical procedure as in the AHP are:

- Weights cannot be determined independently of the units of the single-criterion variables being compared.
- Once the units of the criteria are fixed only *one* normalisation for a set of criteria is admissible.
- Different lower-level sets of normalised sub-criteria preferences should not be combined using upper-level criteria weights because the sets of lower-level priorities are effectively on different scales due to their different normalising constants.

Barzilai thus states that the hierarchical weighted sum aggregation of normalised measures that have been derived using different normalising constants cannot yield meaningful results.

Several other authors have voiced concerns with the AHP normalisation method. Dyer [13] in 1990 changed the normalisation method when attempting to prevent rank reversal:
"The key is to ensure that both the weights on the criteria and the scores of the alternatives on the criteria are normalized with respect to the same range of alternative values."

This would serve to establish a uniform normalising constant for both criteria and alternative preferences resulting in uniform rescaled units for single level models. However, even by doing this the normalisation problem would still be present in multi-level hierarchical preference aggregation models.

Alternatively, if there were only one normalisation constant (i.e. only 1 set of criteria measures on one level) the rescaled units would also be uniform and consistent.

Schenkerman [32] showed in 1991 that rank reversal is caused by eigenvector normalisation. He subsequently showed [33] that the proposed AHP modifications which do prevent rank reversal, work "by undoing normalization of local priorities". Schenkerman [34] recently demonstrated these assertions using a simple geometric estimation problem and showed that the original AHP, the Ideal Mode AHP, and the Pairwise Aggregated approach, all result in an ordering of alternatives that does not really exist. Conversely, he also demonstrates [33] that the Concordant Supermatrix, Referenced AHP, Linking Pins AHP, and the method of Belton and Gear, do give a correct order to the alternatives by undoing normalisation and reducing the method to a simple weighted sum, which is significantly different to the original AHP.

Schoner and Wedley [35] have also identified the normalisation problem, and they have also convincingly demonstrated how normalisation can cause rank reversal due to the change in normalising constant for preferences when new alternatives are added.

3.2 Secondary Problems

3.2.1 Rank Reversal

Rank reversal refers to a changed order of existing alternatives when a new alternative is added. In the early days of the AHP it caused much discussion as to whether it illustrated a fundamental computational weakness or whether it reflected real world decision makers' behaviour and was thus legitimate. Many modifications of the AHP were proposed to prevent rank reversal from happening. However, as discussed previously, several authors have shown that rank reversal is caused by the normalisation of the eigenvectors. Thus, it is a secondary problem and whether or not it does occur among decision makers is a separate and irrelevant question.

3.2.2 Lack of Rating Independency

Axiom 3 states that ratings in a level must not depend on any lower level ratings. In this axiom the meaning of "rating" is somewhat ambiguous since it could be referring to top-down criteria weights, or the rating of the alternatives relative performance against leaf-node criteria. However, since leaf-node criteria are already at the lowest level, this "rating" can only be referring to the top-down criteria relative importance ratings.

Saaty [25, p11] in 1983 stated himself in reference to the matrix of paired comparisons of the relative importance of criteria (here called "attributes" by Saaty):
"Each element (i,j) of the matrix gives the ratio of the average (or total) contribution to cost of attribute i to the average (or total) contribution to cost of attribute j."

In this quote "average" and "total" are assumed to be referring to the set of alternatives contributions (as their priorities at each attribute). This means that the ratings for criteria relative importance are really judging the ratios of the average (or total) utility of the respective criteria with respect to the higher-level criterion. But the contribution of an attribute towards a higher-level criterion is a function of its lower level criteria weights and the alternative ratings for those criteria in hierarchical weighted aggregation. So Saaty himself seems to contradict Axiom 3 in the above quote because of the implicit cross level inter-dependencies in hierarchical weighted aggregation.

Schoner and Wedley [35] have clearly demonstrated, using a simple car selection example, how the relative importance of a criterion (and its computed normalised weight) is proportional to a scaling factor for each criterion, which converts its performance to a common unit (e.g. cost), that is then multiplied by the sum of the absolute values of all alternative performance measures against each criterion. This in essence says the same thing as Saaty above concerning the dependence of criteria relative importance ratings (and thus computed weights) on alternatives' performance ratings against each criterion. And since alternatives are rated always against lowest-level leaf nodes, this must mean upper-level criteria relative importance ratings must be dependent on other lower- level performance ratings for alternatives.

Barzilai [5] has also pointed out the dependency of weight ratings on the units used for alternative performance evaluation per criterion (which is the scaling factor in Schoner and Wedley's equations referred to above). So even if a problem is decomposed into completely independent factors or criteria, and the relative performance of alternatives per leaf-node criterion rated to cancel the individual criteria units out, interdependency still exists between some of the items that are being rated separately. As Barzilai has explained, this is basically caused by Miller's process of hierarchical weighted aggregation of utility or preference. Consequently, Axiom 3 is also questionable.

3.2.3 Cost-Benefit Analysis

Saaty [27] [31] has applied the AHP to cost-benefit analysis by developing global priorities for Cost and Benefits and then dividing them to yield Benefit/Cost ratio priorities. The order of these ratios then yields the order of alternatives based on the Cost per Benefit degree. Two examples of this approach are for comparing the merits of building or not building a road across Sumatra [1], and for deciding whether or not to allow riverboat gambling in one US state [11].

In 1990, Bernhard and Canada [9] raised some doubts about the validity of such Cost-Benefit analysis and demonstrated its limitations:
"Even when benefits and costs are known with certainty and measured in dollars, it is shown that this procedure does not, in general, yield an optimal solution."

And also in [9]:
"More generally, of course, where benefits and costs are not measurable in commensurate units, a more complex analysis would be required. But inevitably, above and beyond benefit and cost output vectors from application of the AHP, that would also continue to require some sort of further consideration and specification of the decision maker's relative willingness to incur various levels of the costs in order to receive corresponding levels of the benefits."

However, the problem those authors are referring to in the second quote is not to do with AHP mechanisms, but the fact that marginal benefit-cost relationships must be considered in any rigorous Cost-Benefit analysis.

But over and above this type of criticism, looms the uncertainty about the validity of the division operation if the input ratings are ordinal measures and the output priority measures for costs and benefits cannot then be ratio scale measures. If one accepts that the scale type places limitations on what operations can legitimately be performed, then the Cost/Benefit ratios would be of limited credibility.

The final problem with AHP Cost/Benefit analysis is that costs and benefits are almost always inter-related and this fact should be addressed by any mathematical method that combines them into an integrated metric. However, such inter-dependency is not addressed in the AHP approach.

4. Summary

The questionable theoretical aspects of the AHP technique that have been highlighted will now be summarised in the order that they are encountered in the application of the technique.

(i) Top-down Rating of the "Relative Importance" of Criteria

It is difficult to know what "relative importance" of criteria means, when comparing two heterogeneous concepts without explicit units of measure in top-down criteria comparisons, and without knowledge of what contributions the respective sub-criteria make.

(ii) The Pairwise Comparison Rating Scale is Ordinal

The ratio comparisons seem to impute a ratio scale to the ratings and produce absolute measures by cancelling out units for criteria. However, this is not the case since the linguistic or numerical measures applied are on ordinal scales. So A/B = 5 cannot mean A = 5B unless units are assigned. Thus, any numbers assigned are necessarily ordinal measures, and this implies that the eigenvalue polynomial computation is inadmissible.

(iii) The Eigenvalue Method for Determining Priorities.

There seems to be no valid reason why the right eigenvector method does balance out inconsistent ratings, especially since left and right eigenvectors may yield different results. This uncertainty is in addition to whether or not the eigenvalue computation is admissible by scale type limitations.

(iv) The Normalisation Problem

Normalisation of the weight and alternative preference vectors causes anomalies in both single level and multi-level hierarchical aggregation of priorities, and is one of the reasons for rank reversal.

(v) Additive Aggregation of Priorities

For additive aggregation all criteria must be independent and not inter-related, which is often not the case.

5. Conclusions

This technical note has summarised various critical analyses of the AHP that have occurred over the last 20 years or so. The feature that initially sparked many investigations was rank reversal and this caused much discussion about whether it was legitimate or not. Regardless of whether it can occur with real world decision makers, it has been convincingly shown to be a function of normalisation. Consequently, it is considered to be a secondary category problem in this report. In contrast, the more fundamental primary category of problems has been defined and the problems identified within this category are: scale misinterpretation, comparison matrix eigenvalue evaluation, and multiple normalisations in hierarchical aggregation of priorities. Moreover, it has been shown that the axiomatic foundations of AHP are also questionable.

In general, it is not possible to validate decision analysis techniques based on subjective scoring such as the AHP when they are applied to strategic decisions with abstract criteria. This fact has resulted in the AHP being used in a wide variety of applications, which in turn has established the method with a sort of *de facto* credibility. In their enthusiasm to apply the AHP with its very user-friendly software, analysts not infrequently construct models that also violate the most basic constraint of independence of criteria or factors. The increased complexity of procedures for aggregating inter-dependent information may be partially responsible for this, plus the fact that such methods are scarce. However, Saaty has proposed another technique, called the Analytic Network Process to be applied when independence of criteria does not exist. Unfortunately the scale misinterpretation problem is again present so its results also are very questionable. And besides that there are further higher-level assumptions and procedures that are also questionable.

It is curious that the large amount of literature focusing on comparing different AHP computational mechanisms is largely inconclusive, and all tacitly seem to accept that the input ratings are actually ratio scale measures that allow complicated algebraic operations to be validly performed. Despite numerous claims by the AHP school that the method gains it rigour because it uses ratio scale measures, it is obvious that there is a fundamental misunderstanding in what the different types of scale mean. As Stevens has pointed-out, this is a common problem in scientific literature because it has primarily been concerned with matters of the physical realm. And because no bells and whistles sound when inadmissible operations are performed on measures, sophisticated computational mechanisms can effortlessly be applied which convinces others of the method's validity by virtue of their "sophistication". The unfortunate conclusion is that the many simulations of computational AHP refinements are all meaningless because they also perform inadmissible operations. Although some comparisons of quantitative factors may invoke a quasi-ratio scale rating of some measurable property, Saaty's proposed scale is not a true ratio scale because of its lower and upper limits, and the absence of an absolute zero. And needless to say, comparisons of qualitative factors cannot yield ratio scale measures.

Overall, even without the ordinal scale problem, there are enough questionable features in the AHP to severely doubt the validity of the output priorities. With this in mind, the method should be applied with great caution. It should also be noted that it is not only the AHP that is subject to some of these criticisms and several other techniques in the field of multi-criteria or multi-attribute decision analysis also have similar limitations. At the present time we are examining several other decision analytic techniques that have been proposed recently and which attempt to avoid the pitfalls described. Needless to say, if decision analytic methods are being applied to make important Defence decisions, the method applied should be theoretically sound. Only then can there be confidence in the analytical results.

6. References

1. Azis, I. [1990] AHP in the Benefit-Cost framework: A post-evaluation of the trans-Sumatra highway project, *European J. Operational Research*, **48**, pp. 38-48.
2. Barzilai, J., Cook, W. and Golany, B. [1987] Consistent weights for judgments in matrices of relative importance of alternatives, *Operations Research Letters*, **6** (3), pp. 131-134.
3. Barzilai, J. [1997] Deriving weights from pairwise comparison matrices, *J. Operational Research Society*, **48** (12), pp.1226-1232.
4. Barzilai, J [1998] On the decomposition of value functions, *Operations Research Letters,* **22**, pp.159-170.
5. Barzilai, J [1998] Understanding hierarchical processes, *Proc. 19th American Society for Engineering Management*, pp. 1-6.
6. Barzilai, J. [1999] On MAUT, AHP, and PFM, *Proc. 5th Int. Symposium on the AHP*, Kobe Aug12-14, 1999. pp. 57-60.
7. Barzilai, J. [2000] Theory of measurement for voltage and affine variables, *Proc. IEEE Int. Conf. Systems, Man and Cybernetics*, Oct 8-11, Nashville, pp. 636-640.
8. Barzilai, J. [2001] Notes on the Analytic Hierarchy Process, *Proc. NSF Design and Manufacturing Research Conference*, Tampa 2001, pp. 1-6.
9. Bernhard, R. and Canada, J. [1990] Some problems in using Benefit/Cost ratios with the AHP, *The Engineering Economist,* **36** (1), pp. 56-66.
10. Churchman, C. and Ackoff, R. [1954] An approximate measure of value, *Operations Research*, **2** (2), pp. 172-181.
11. Clayton, W., Wright, M. and Snodgrass Sarver, W. [1993] Benefit Cost analysis of riverboat gambling, *Math. Compt. Modelling*, **17** (4/5), pp. 187-194.
12. Crawford, G. and Williams, C. [1985] A note on the analysis of subjective judgment matrices, *J. Mathematical Psychology*, **29**, pp.387-405.
13. Dyer, J. [1990] Remarks on the Analytic Hierarchy Process, *Management Science*, **36** (3), pp. 249-258.
14. Forman, E. [1993] Facts and fictions about the AHP, *Math. Comput. Modelling*, **17** (4/5), pp. 19-26.
15. Forman, E. and Gass, S. [2001] The Analytic Hierarchy Process- An Exposition, *Operations Research*, **49** (4), pp.469-486.
16. Golany, B, and Kress, M. [1993] A multicriteria evaluation of methods for obtaining weights from ratio-scale matrices, *European J. Operational Research*, **69**, pp.210-220.
17. Harker, P. and Vargas, L. [1987] The theory of ratio scale estimation: Saaty's Analytic Hierarchy Process, *Management Science*, **33** (11) pp.1383-1403.
18. Jaiswal, N. [1997] *Military Operations Research* (Chapter 8), Dordrecht: Kluwer, pp. 216-224.
19. Luce, R., Krantz, D., Suppes, P. and Tversky, A. [1990] *Foundations of Measurement* (Vol. 3), New York: Academic Press.

20. Mikhailor, L. and Singh, M. [1999] Comparison analysis of methods for deriving priorities in the AHP, *Proc. IEEE Systems, Man and Cybernetics Symposium* (Vol.1), pp. 1037-1042.
21. Miller, J. [1966] *The Assessment of Worth: A Systematic Procedure and Its Experimental Validation*, Doctoral Dissertation, MIT.
22. Roberts, F. [1979] *Measurement Theory*, New York: Addison-Wesley..
23. Saaty, T. [1977] A scaling method for priorities in hierarchical structures, *J. Mathematical Psychology*, **15**, pp. 234-281.
24. Saaty, T. [1980] *The Analytic Hierarchy Process*, New York: McGraw-Hill.
25. Saaty, T., Vargas, L., and Wendell, R. [1983] Assessing attribute weights by ratios, *Omega*, **11**(1), pp. 9-12.
26. Saaty, T. and Vargas, L. [1984] Inconsistency and rank preservation, *J. Mathematical Psychology*, **28**, pp. 205-214.
27. Saaty, T. [1985] *Analytical Planning*, New York: Pergamon Press.
28. Saaty, T. [1986] Axiomatic foundations of the AHP, *Management Science*, **32**, pp.841-855.
29. Saaty, T. [1993] What is relative measurement? The ratio scale phantom, *Mathematical Computer Modeling*, **17** (4/5), pp.1-12.
30. Saaty, T. [1996] *Decision Making with Dependence and Feedback: The Analytic Network Process*, Pittsburg: RWS Publications.
31. Saaty, T. and Vargas, L. [2001] *Models, Methods, Concepts and Applications of the Analytic Hierarchy Process*, Boston: Kluwer.
32. Schenkerman, S. [1991] The cause of rank reversal in weighted criteria decision support models, *Proc. 20th Annual Northeast Conference of the Decision Sciences Institute*.
33. Schenkerman, S. [1994] Avoiding rank reversal in AHP decision-support models, *European J. Operational Research*, **74**, pp. 407-469.
34. Schenkerman, S. [1997] Inducement of non-existent order by the AHP, *Decision Sciences*, **28** (2), pp. 475-482.
35. Schoner, B. and Wedley, W. [1989] Ambiguous criteria weights in AHP: consequences and solutions, *Decision Sciences*, **20**, pp. 462-474.
36. Schoner, B., Wedley, W. and Choo E. [1992] A rejoinder to Forman on AHP, with emphasis on the requirements of composite ratio scales, *Decision Sciences,* **23**, pp. 509-517.
37. Scott, D. and Suppes, P. [1958] Foundational aspects of theories of measurement, *J. Symbolic Logic*, **23** (2), pp. 113-128.
38. Stevens, S. [1946] On the theory of scales of measurement, *Science*, **103**, pp. 607-680.
39. Stewart, T. [1992] A critical survey on the status of multiple criteria decision making theory and practice, *OMEGA*, **20** (5/6), pp. 569-586.
40. Takeda, E., Cogger, K. and Yu, P. [1987] Estimating criterion weights using eigenvectors: A comparative study, *European J. Operational Research*, **29**, pp. 360-369.

DEFENCE SCIENCE AND TECHNOLOGY ORGANISATION
DOCUMENT CONTROL DATA

1. PRIVACY MARKING/CAVEAT (OF DOCUMENT)

2. TITLE Uncertainties in the Analytic Hierarchy Process	3. SECURITY CLASSIFICATION (FOR UNCLASSIFIED REPORTS THAT ARE LIMITED RELEASE USE (L) NEXT TO DOCUMENT CLASSIFICATION) Document (U) Title (U) Abstract (U)
4. AUTHOR(S) Lewis Warren	5. CORPORATE AUTHOR Information Sciences Laboratory PO Box 1500 Edinburgh South Australia 5111 Australia

6a. DSTO NUMBER	6b. AR NUMBER	6c. TYPE OF REPORT	7. DOCUMENT DATE
DSTO-TN-0597	AR-013-275	Technical Note	December 2004

8. FILE NUMBER	9. TASK NUMBER	10. TASK SPONSOR	11. NO. OF PAGES	12. NO. OF REFERENCES
N9505/23/57	JTW 02/304	HSOD	22	40

13. URL on the World Wide Web	14. RELEASE AUTHORITY
http://www.dsto.defence.gov.au/corporate/reports/DSTO-TN-0597.pdf	Chief, Command and Control Division

15. SECONDARY RELEASE STATEMENT OF THIS DOCUMENT

Approved for public release

OVERSEAS ENQUIRIES OUTSIDE STATED LIMITATIONS SHOULD BE REFERRED THROUGH DOCUMENT EXCHANGE, PO BOX 1500, EDINBURGH, SA 5111

16. DELIBERATE ANNOUNCEMENT

No Limitations

17. CITATION IN OTHER DOCUMENTS Yes

18. DEFTEST DESCRIPTORS

Decision making
Heuristics

19. ABSTRACT
When choosing a decision analysis technique to model a particular complex decision the fundamentals of the technique chosen should be understood by the analyst, and they should be appropriate for the characteristics of the problem itself. For analysing such complex decisions the Analytic Hierarchy Process is one of the most commonly used techniques. This technical note highlights a number of theoretical issues, some well-known and others less well-known, that introduce a considerable degree of uncertainty into the computed output priorities for the decision alternatives.