J. Opl Res. Soc. Vol. 41, No. 11, pp. 1073–1076, 1990
Printed in Great Britain.

0160-5682/90 $3.50 + 0.00

Some Comments on the Analytic Hierarchy Process

R. D. HOLDER
Consultancy Group, SD-Scicon UK Limited, London

Some comments are made on the analytic hierarchy process of Thomas L. Saaty, which has gained widespread acceptance as a valuable tool for multicriteria decision-making. Saaty's validation of the method against physical laws is criticized, a multiplicative scale is suggested for making judgements, and the problem of rank reversal is discussed with reference to two published papers.

Key words: analytic hierarchy process, multicriteria decision analysis

INTRODUCTION

The analytic hierarchy process (AHP) of Thomas L. Saaty has emerged in the last 10–15 years as a major tool in multicriteria decision analysis. The method has been well documented by Saaty himself,[1–3] and case studies using it have been reported by many authors, including some recently in the *Journal of the Operational Research Society*.[4,5]

During the course of examining and using the AHP, we have noted certain inconsistencies in the method and its application which we feel ought to be brought to the attention of other practitioners. These inconsistencies are the subject of this note. We believe that only a version of the method modified to be internally consistent should be used in practice.

VALIDATION OF THE AHP USING PHYSICAL DATA

Saaty[3] claims that judgements elicited and weights derived using the subjective AHP can be validated from the laws of physics for appropriate applications. This claim needs to be treated with some caution, as the following example, given by Saaty,[3] illustrates.

In an experiment, four identical chairs were placed at various distances from a floodlight. Some children were placed beside the light and asked to give pairwise comparative judgements about the relative brightness of the chairs, using Saaty's semantic scale. These comparisons were fed into the AHP to give overall weights for the brightness of the chairs. Saaty then goes on to say, 'the inverse square law of optics is now used to test these judgments'. Indeed, it turns out that the weights are very close to the normalized inverse squares of the distances of the chairs from the light.

Now, as the experiment is described, it is apparent that there is a two-way path of light from the floodlight to the chairs and back again. This gives rise, not to an inverse square law, but probably to something nearer an inverse fourth-power law of degradation of intensity. Moreover, the response of the human eye is unlikely to be linear in intensity.

There is also the further point that, in this experiment, the distances of the chairs from the floodlight were chosen very fortuitously. Thus the maximum ratio of distances was close to 3, so that the 1–9 scale of comparisons would just be able to cope with an inverse square law, had that applied.

Validation against the laws of physics thus appears at best rather dubious, and we are left with the AHP giving us little more than a correct rank ordering according to brightness—no great achievement.

THE JUDGEMENT SCALE

We believe that the 1–9 judgement scale advocated by Saaty is problematical in a number of ways.

When making pairwise comparisons using the AHP, the decision-maker (DM) is presented only with English language descriptions of relative importance. Thus, it is hidden from him that what he is really being asked to do is to estimate ratios of weights for pairs of criteria. For this the linear 1–9 scale seems singularly inappropriate; indeed, it is not clear why the user should not be asked for numerical estimates direct and given a free hand as to the numbers he can choose.

On the question of consistency, which is the *raison d'être* of the AHP, again what this means for the user is concealed and, moreover, it is impossible for him to be consistent if he wants to be.

A perfectly consistent pairwise comparison matrix a_{ij} has the transitivity property that

$$a_{ij}\,a_{jk} = a_{ik}$$

for all i, j, k. There would seem to be no rationale for the corresponding logic of the semantic scale—for example,

> A is weakly more important than B (3 on Saaty's scale) and B is weakly more important than C (3 on Saaty's scale) imply that A is absolutely more important than C (9 on Saaty's scale)—

especially in view of the gaps in the scale. Indeed, the above logic does violence to the normal usage of the English language.

With the linear scale, the user cannot be consistent because the scale is not *complete*. He may well want to say that A is twice as important as B, and A is 3 times as important as C, and B is $1\frac{1}{2}$ times as important as C, yet he is constrained to make the last judgement 1 or 2. Of course, the scale is further incomplete and unnecessarily restricting because of the arbitrary cut-off at 9 for the maximum allowable ratio of weights.

Equating the adjectival descriptions with a multiplicative scale would overcome the first of the above problems of completeness. Elicitation from the DM, or the DM in consultation with the OR analyst, of the maximum estimated ratio of weights would also solve the second problem. Such a scale would have the form

$$1, \alpha, \alpha^2, \alpha^3, \ldots, \alpha^n,$$

where α represents the smallest detectable ratio of weights and α^n the largest estimated ratio for the problem in hand.

Semantic equivalents could be given for the multiplicative scale just as Saaty does for the linear 1–9 scale. For an example in which the scale had the same number of points as Saaty's, indeed precisely the same semantics could be used. Weak importance would be represented as α^2, essential importance as α^4, etc. up to absolute importance as α^8, and now, since $\alpha^2 \times \alpha^2 = \alpha^4$, two weakly more importants would combine to give only an essentially more important. However, it makes more sense, as suggested above, to leave the minimum discernible ratio (α) or the maximum estimated ratio (α^n) to be chosen by the DM for the particular application in hand.

The reader is referred to the paper by Lootsma,[6] where similar ideas for a geometric (multiplicative) scale, also using pairwise comparisons, are applied to the area of conflict resolution by cost-benefit trade-off analysis.

RANK REVERSAL

Given a straightforward application of the AHP in Saaty's original form, the fallacy of rank reversal of candidates may occur if either a new candidate is introduced or an existing candidate is removed. The problem was first reported by Belton and Gear,[7] and has since been widely discussed in the literature, most recently by Schoner and Wedley.[8] However, the message has clearly not been driven home in the AHP users' community since papers continue to be published using what can only be described as a flawed method.

Sinuany-Stern[4] used the AHP in its original form to rank 16 Israeli football teams against six criteria. With so many teams, it is difficult to demonstrate rank reversal because the impact of the introduction of an extra team, or the removal of an existing one, is diluted when normalization of the score vectors is carried out. Nevertheless, it can be shown that the introduction of a further three identical, hypothetical teams which rank equal best with team 2 on the 'players' criterion and equal worst with the existing worst performers on all other criteria leads to rank reversal of teams 6 and 13. That is to say, these teams were originally placed third and fourth overall, and are now placed fourth and third respectively. In the calculation it was assumed that pairwise comparisons, and hence ratios of scores, remained constant for existing teams. The rank reversal is thus a quirk of the method, and not a real, expected effect.

A more easily demonstrable example of rank reversal occurs in the paper by Roper-Lowe and Sharp,[5] in which the application is to the choice of a computer operating system for British Airways. This example is the more serious because these authors claim to know about the problem and to have implemented a version of AHP free from it due to Belton and Gear.[7] What these authors do is to normalize score vectors so that the best performing candidate has a score of unity on a given criterion. If the third of their three candidates (called ACP) is removed, then, on criteria where ACP is best, renormalization leads to a change in scores for the first two candidates. The new table is given here as Table 1. It has been constructed on the reasonable assumption, made for lack of information on the actual pairwise comparison matrices, that all judgements were perfectly consistent, and therefore that the ratios of scores are preserved.

TABLE 1. *Option scores and weights of bottom criteria for Roper-Lowe and Sharp application of AHP with third candidate removed*

Scores	DB2	TPF	Weights
Flexibility	1	0.23	0.190
Tactical development	0.24	1	0.008
Database integrity	1	0.15	0.095
Response times	0.27	1	0.024
Programmer productivity	1	0.24	0.075
Future BA expertise	1	0.45	0.025
Programmer availability	1	0.26	0.011
Cutover risk	0.30	1	0.103
Functional risk	0.20	1	0.297
Support risk	1	1	0.035
BA expertise	0.18	1	0.028
Database changes	0.19	1	0.005
Communications changes	0.24	1	0.005
Network up	0.91	1	0.078
Message integrity	1	0.23	0.016
Total of weight × score	0.62	0.68	

The rank reversal of options DB2 and TPF which has now occurred is precisely what would have happened had the procedure described in the paper been adopted for just the first two candidates. Again, it is a product of a faulty method rather than a real effect.

The problem of rank reversal arises because of the insistence that score vectors are normalized, either so that components sum to unity or so that the largest component is unity, before composition with weights, and because weights are elicited without reference to scales for performance against criteria. Roper-Lowe and Sharp are aware of this, yet in practice have elicited comparisons of criteria independently (not necessarily even by the same people) of the comparisons of candidates against the criteria.

If credibility of results is to be achieved, it is essential that a method be used which is not subject to this inconsistency. This can be achieved in a number of ways—for example, by having candidates' performance in mind when the weighting of criteria is done, in which case the weights would be rederived on the introduction of new, or omission of existing, candidates. Schoner and Wedley[8] show how both conventional AHP and the Belton-Gear approach might be adapted in this way. An alternative would be to abandon pairwise comparison of candidates altogether at the bottom level of the hierarchy and simply score candidates using objective scales.

ACCURACY OF RESULTS

Normal usage of AHP takes judgements on the semantic scale and produces results on a numerical scale, as in the previous section. It is to the credit of Roper-Lowe and Sharp that they recognized that, since DB2 came out only marginally ahead of TPF in their paper, no firm conclusions could be drawn. Perhaps a further suggestion is that results could be reinterpreted on the semantic scale by dividing the final scores of pairs of candidates. If this is done for the best (DB2) and worst (ACP) candidates, we obtain a ratio of 1.18, not nearly 'weakly' different on the Saaty semantic scale used.

CONCLUSIONS

We have presented a number of criticisms of the AHP and its use in practice:

(i) The AHP is not as well validated as might be supposed, judging from Saaty's illuminated chairs example.

(ii) The scale used for comparisons is illogical: if a scale is used, it ought to be multiplicative and not linear.

(iii) The method must be modified to avoid the problem of rank reversal.

(iv) It makes sense to reinterpret numerical results in terms of the original semantic scale.

It is clear that there is much room for more empirical work on scales and the like; nevertheless, in the meantime we would urge practitioners to use a modified version of the AHP which is at least internally consistent.

Acknowledgements—The author wishes to thank his colleagues at SD-Scicon UK Limited for many useful discussions on this subject, and also the Ministry of Defence (UK) for a contract sponsoring related work in this area.

REFERENCES

1. T. L. Saaty (1977) A scaling method for priorities in hierarchical structures. *J. Math. Psychol.* **15,** 234–281.
2. T. L. Saaty (1980) *The Analytic Hierarchy Process.* McGraw-Hill, New York.
3. T. L. Saaty (1983) Priority setting in complex problems. *IEEE Transactions on Engineering Management* **EM-30,** 140–155.
4. Z. Sinuany-Stern (1988) Ranking of sports teams via the AHP. *J. Opl Res. Soc.* **39,** 661–667.
5. G. C. Roper-Lowe and J. A. Sharp (1990) The analytic hierarchy process and its application to an information technology decision. *J. Opl Res. Soc.* **41,** 49–59.
6. F. A. Lootsma (1989) Conflict resolution via pairwise comparison of concessions. *Eur. J. Opl Res.* **40,** 109–116.
7. V. Belton and T. Gear (1983) On a shortcoming of Saaty's method of analytic hierarchies. *Omega* **11,** 228–230.
8. B. Schoner and W. C. Wedley (1989) Ambiguous criteria weights in AHP: consequences and solutions. *Decis. Sci.* **20,** 462–475.

LINKED CITATIONS

You have printed the following article:

Some Comments on the Analytic Hierarchy Process
R. D. Holder
The Journal of the Operational Research Society, Vol. 41, No. 11. (Nov., 1990), pp. 1073-1076.
Stable URL:
http://links.jstor.org/sici?sici=0160-5682%28199011%2941%3A11%3C1073%3ASCOTAH%3E2.0.CO%3B2-K

References

[4] **Ranking of Sports Teams via the AHP**
Zilla Sinuany-Stern
The Journal of the Operational Research Society, Vol. 39, No. 7. (Jul., 1988), pp. 661-667.
Stable URL:
http://links.jstor.org/sici?sici=0160-5682%28198807%2939%3A7%3C661%3AROSTVT%3E2.0.CO%3B2-S

[5] **The Analytic Hierarchy Process and Its Application to an Information Technology Decision**
G. C. Roper-Lowe; J. A. Sharp
The Journal of the Operational Research Society, Vol. 41, No. 1. (Jan., 1990), pp. 49-59.
Stable URL:
http://links.jstor.org/sici?sici=0160-5682%28199001%2941%3A1%3C49%3ATAHPAI%3E2.0.CO%3B2-I

NOTE: *The reference numbering from the original has been maintained in this citation list.*