FOREWORD

THE AHP CAPTURES ORDER SO ESSENTIAL IN MAKING DECISIONS

Thomas L. Saaty
University of Pittsburgh

Many decision problems are hierarchic in nature in that they have goals, criteria, subcriteria and so on down to the alternatives of choice. A hierarchic structure by the natural stratification of its levels gives one a good start in the quest for order. Generally, the objective of decision theory is to order the alternatives at the lowest level of the hierarchy. A weak form of this order is to "totally order" or line up the alternatives sequentially according to overall preference on multiple criteria. A strong version is to map the alternatives of which there are $m > 0$ to the interval $[0,1]^m$, to represent order on a numerical scale according to importance or something similar (called rank or ranking). The components representing the priorities of the alternatives belong to the same ratio scale and sum to unity. This is what one does in the Analytic Hierarchy Process (AHP).

Decision problems are unstructured and need to be structured to derive an optimum ranking. For most problems, the hierarchy and relations within it are only partly known to a decision maker. We need a mathematical formalism to assist him to develop rank from incomplete information. The structure is at least as important as the ranking procedure used to derive the order within the structure. We already know some general structural principles for organizing a decision problem as a hierarchy. We also know of ways to structure a decision problem from the general to the particular partly learned through the pairwise comparison procedures we use in the AHP, or as a discrete system of components with feedback including dependence within and between the components. Finally we have studied decision problems in neural network structures on manifolds with dependence and feedback. We can use the formal theory of the AHP in all three of these structures.

There are two legitimate concerns of decision theory. The first is how to derive an optimum rank for the alternatives of a multicriteria decision with respect to each criterion and also an optimum overall rank with respect to all the criteria, and the second is how to interpret rank preservation and reversal in closed and in open sequential decision structures. I shall confine my discussion to decision problems represented by hierarchic structures, sometimes viewed as a very particular case of discrete feedback systems. These two cases involve the use of positive reciprocal matrix operators whereas the third case of a manifold extends these ideas to positive operators of the Fredholm type with reciprocal kernels. My purpose is to give an overview and open up for inspection the philosophy underlying decision making with the AHP.

Paired Comparisons and the Derived Order

We need to make a distinction between the mathematics of decision problems and that of physics and engineering around which much of our understanding of the physical world has been developed. In the physical world, topology and in particular metric spaces, are useful to study limiting operations and the concepts of nearness and good approximation. But when we rank elements in a hierarchy it is not enough to consider the metric idea of closeness, we also need a best way to derive rank that embodies the numerical preferences or dominance relations expressed in the judgments. It is possible to introduce different metrics (e.g. least squares, logarithmic least squares) on a space and obtain approximations for rank but the orders may be different, with the top alternatives and their ranks depending on the metric chosen. Topological closeness need not signify order closeness. In ordering three elements, the first and the third may be judged closer by a metric than the first and the second which lies between them. Thus the question is, if we use an infinite process to obtain a limiting rank, what sort of metric do we need to generate a unique order from judgments? In decision making we must generate a unique rank for the alternatives according to a special metric with which the idea of closeness in n-dimensional space carries with it preservation of rank. Here small perturbations in judgments must lead to small changes in priority and little or no change in rank as appropriate.

A hierarchy of three levels: goal, criteria, alternatives is what multicriteria people often talk about. Most decisions go beyond a three level hierarchy though it is neither the simplest nor the most complex type of decision structure. Let us look at the order generated in the simplest decision structure, a two level hierarchy consisting of a criterion, and several alternatives to be ordered on that criterion or attribute. In the AHP we obtain this order as the principal eigenvector of a paired comparisons matrix of judgments. In these paired comparisons, dominance is expressed in the form of ratios and their reciprocals, with respect to an attribute. Here one examines two alternatives and asks : on the whole or on the average are the two equal in possessing the attribute? If not, take the one which possesses the attribute less than the other and determine how many times more the larger alternatives possess the attribute than the smaller one. The latter then serves as the unit of the comparison and the other is assigned a real number as an absolute multiple of the unit, often approximated by an integer. If we anticipate deriving a ratio scale from the judgment with values w_i for alternative i and w_j for alternative j, then the ratio is $(w_i/w_j)/1$ and its reciprocal $1/(w_i/w_j)$ in the transpose position. Thus we derive a ratio scale from an absolute scale of numbers.

We assume that in structuring a decision problem, alternatives are grouped in homogeneous clusters so that one essentially compares one alternative with another that is a small perturbation of it. If an element is much larger than the rest, it is removed from that cluster and placed in another cluster of elements similar to it. By sharing an element, measurement on the two clusters can be unified. The reason for this clustering is that people can only deal with comparisons of homogeneous objects and cannot compare widely disparate elements. Clustering assists judgment elicitation and improves the accuracy of the final answer.

Returning to our matrix of paired comparisons, we see that the elements being compared must be sufficiently close to be able to relate them with accuracy. The need to cope with human perceptual limitations is the main reason why the fundamental scale of the AHP is confined to the absolute numbers 1, 2, ..., 9 which correspond closely to a verbal scale that also spans our ability to make distinctions. The scale is only an indicator of order of magnitude and not a rule to be followed to the letter. If one has precise decimals, one can use them instead. Note that this process of translating from words to numbers does not come out of previous theories, it is simply assumed first and afterward tested in a large number of cases. In practice it works well. This very scale contains as subsets the geometric progressions 1, 2, 2^2, 2^3 and 1, 3, 3^2, thought by some to be preferred to the 1-9 scale. It is difficult to assign larger values from whatever scale because people are not capable of making widely disparate comparisons. If someone can use a wider scale with accuracy, let them. To deal with far out alternatives we need to use clustering and should get the same result without loss of generality. Too narrow a scale would be limiting and inefficient, and too wide a scale would sacrifice precision. Our goal is to try to be both valid and consistent.

Let us make a useful observation about attaching numbers to stimuli. Though we look for ways to represent our mental response to stimuli, we are not always able to respond to them with the same intensity they actually occur in the real world, as measured by instruments on the few scales we have. It is how we interpret the world based on information processed by us that matters in decision making. The fact that there are very large measurements of distance and weight on our instruments or estimated mathematically, does not mean that we can in fact sense and attach a meaning to different readings or values. Measurements are representations of stimuli which we need to interpret just as we do stimuli we receive directly like light and sound. Some stimuli are intense and some are weak. We must interpret them for meaning when they go beyond our thresholds of perception.

Some people have improvised logarithmic or power laws for our response on entire ranges of phenomena. In the AHP because our approach is constructive, rather than descriptive, we can deal with them in contiguous homogeneous ranges measured in relative terms and do not need a law that applies uniformly (and none does) across the full range. Sometimes measurements are meaningless. If you have any doubt about it you can ask what in your mind is the difference between one trillion and one trillion one million kilometers. We each have our own ranges of significance of numbers for each kind and range of measurement. Note that the significance of the million is lost in the very large and unfathomable trillion.

In doing paired comparisons, we note that the idea of consistency is important. One can force the judgments to be consistent but may cause a loss in accuracy. When judgments are consistent all methods people have suggested to rank the alternatives: taking the average over the rows, multiplying the elements and taking their n^{th} root, using the method of least squares, lead to the same answer and there is no reason to prefer one over the others.

We note that dominance in this case is transitive. In other words if A is preferred 3 times over B, and if A is preferred 5 times over C, and if C is then preferred 3/5 times over B we still have from A to C to B that A is preferred 3 times to B. In other words when there is consistency, A is preferred to B the same whether it is directly compared with it or indirectly through C. In general, we have $A^k = n^{k-1}A$, and it is enough to use the matrix A to describe all the relations between the elements. As I said, it does not matter what method one uses to derive the scale because all yield the same answer.

The question now is, what if one does not feel sure about taking a minimum of judgments but must make all the comparisons, in other words, what if the consistency condition $a_{ij}\, a_{jk} = a_{ik}$ is no longer satisfied? Such an outcome is likely to be the case particularly when one is dealing with intangible criteria.

Now in the consistent case the solution (which coincides with that obtained by all methods of optimization) is derived from:

$$Aw = nw.$$

where n is the order of A. A perturbation theorem assures us that we need to solve the corresponding problem $Aw = \lambda_{max}w$, where λ_{max} is the largest or principal eigenvalue of A, known to always exist when A is positive. The solution of this problem is not a matter of choice, it is always there with its ranking. It is thrust upon us without making any new assumptions.

The next idea is to show that the principal eigenvector is the way to capture dominance in the judgments when the matrix is inconsistent. With inconsistency, it is no longer true that $A^k = n^{k-1}A$ and we must consider all possible relations of dominance which means that we must account for dominance along paths whose lengths are 1, 2, 3,... in A, A^2, A^3, ... respectively. We can prove that the mean dominance from all these matrices converges to the principal eigenvector of A known to always belong to a ratio scale. In this manner the eigenvector captures dominance.

The theorem of Perron which says that if A is a positive linear transformation on R^m, then there is an $x_o > 0$ such that for all $x \geq 0$, $A^n x$ converges in direction to x_o so that

$$\frac{A^n x}{||A^n x||} \rightarrow \frac{x_0}{||x_0||}$$

According to Birkhoff [1, 2], this theorem is a special case of the contraction mapping theorem which says that if A is a contraction on a complete metric space (X,D) mapping X into X, i.e., for some k<1, D(Ax, Ay) ≤ kD(x, y) for all x,y ∈ X, then there exists $x_o \in X$ such that $A^n x \rightarrow x_o$, for all x∈X.

Birkhoff observed that there is a metric D on x in which all positive linear transformations acting on the set of rays X in R^m_+ satisfy the contraction condition because convergence in rays is also convergence in direction. The unique metric D, invented by Hilbert for non-Euclidean geometries specialized to R^m_+ that makes a positive (or even a nonnegative) matrix into a contraction that satisfies the theorem, is given by

$$D(x,y) = \log \frac{\max_i (x_i/y_i)}{\min_i (x_i/y_i)}$$

where x_i and y_i are the i^{th} coordinates of the vectors x and y. It is a metric on rays because D(ax,by) = D(x,y) for a,b > 0 and thus it is order preserving. According to Kohlberg and Pratt [2], a method that needs to preserve order must converge along rays and thus cannot use an arbitrary metric.

The AHP also has a nice index based on the structural parameter λ_{max} that captures all the inconsistencies in a single index λ_{max}-n. It can be used to decide whether the judgments are sufficiently consistent to justify deriving a valid ranking, or whether the decision should be made at all because of loose information. Sometimes one hears it asked: Why use paired comparisons and not some other form of comparisons to express the judgments? A reply to that is that mathematically pairwise comparisons is the only way we have to derive the relative priorities in the form of the principal eigenvector that captures the dominance in the judgments.

Extending Rank Order to a Hierarchy

It has been shown [3] that the principal eigenvector represents the relative dominance or rank of each element in the paired comparison matrix A. This dominance is captured by the principal eigenvector. When the judgments are consistent each power of A is a constant multiple of A and hence it is sufficient to normalize the row sums of A to derive the principal eigenvector which gives the relative priorities or ranks of the elements. When A is inconsistent, one must consider the mean of corresponding components obtained as the normalized row sums of every power of A. the result is again the principal eigenvector of A.

Now let us turn to the question of deriving an optimum overall rank in a multiattribute process. Decisions are sometimes one shot affairs and sometimes sequential and we need to understand both. A hierarchy is a decomposition into levels and elements in levels of a complex dominance situation. In a closed hierarchy, the one shot affair, all the elements of the structure are assumed to be included at the start. The elements in each level of the hierarchy are mutually independent with respect to shared attributes in the immediately preceding level on which they depend and with respect to which they are ranked according to dominance, but they are dependent with respect to relative measurement. They are also assumed to be independent of the elements in lower levels of the hierarchy. Thus the rank of the elements with respect to each element in an adjacent upper level is obtained as the principal eigenvector of their paired comparison matrix. Two alternatives are said to be mutually independent if the intensity with which an alternative possesses an attribute is not influenced by the existence of the other alternative and conversely.

The overall ranks of the alternatives in a decision are derived through a generalization of the principal eigenvector of a single pairwise comparison matrix. We refer to the method of deriving these ranks as the Hierarchic Composition Principle. Here we construct a general matrix to represent the relative dominance of the elements in the entire hierarchy. It incorporates all the separate eigenvectors derived throughout the hierarchy. Such a matrix is known as a supermatrix A. The supermatrix has zero submatrices in those positions corresponding to non-interacting levels and an identity submatrix in the last row and column positions corresponding to the alternatives because that level is an absorbing state. Two levels are said to interact if one is dependent on the other or conversely.

In graph theoretic terms, a hierarchy is a directed graph from the goal to the bottom level of alternatives. The levels of the hierarchy are the vertices of a path that connects the top goal to the level of alternatives. In addition, each of these vertices is itself a strongly connected subgraph. The directed graph gives rise to the supermatrix which represents the dominance-reachability of any vertex from any other vertex of the hierarchy. As in the simpler paired comparisons representation of dominance, the limiting power of this reachability matrix yields the overall dominance of each element in the hierarchy including the lowest level alternatives.

When the hierarchy is closed, the limit of the supermatrix exists because it is column stochastic, irreducible, and imprimitive. In fact this limit is equal to the matrix raised to a power, one less than its order. We have:

<u>Theorem</u> (Hierarchic Composition Principle for closed systems) : Given a closed hierarchy H and its corresponding supermatrix A, the ranks of the alternatives at the bottom level of H with respect to the goal are uniquely given by the (n,1) entry of $\lim_{k \to \infty} A^k$.

When the hierarchy is open allowing for sequentially adding elements but particularly adding alternatives there are two cases. The first is the normative case. Here in the supermatrix each alternative is assigned a value 1 under the appropriate intensity for each criterion and a value of zero under the remaining intensities. The second is the descriptive case where in the supermatrix the priorities of the alternatives are adjusted with respect to an ideal alternative.

<u>Corollary</u> (Hierarchic Composition Principle for Open Systems) : Given a hierarchy H with an open level of alternatives and supermatrix A of order n, the ranks of the alternatives are uniquely given by the (n,1) entry of A^{n-1}.

This entire procedure gives an optimum final order. Any other metric procedure which derives the optimum order from consistent judgment matrices must also use the Hierarchic composition Principle for synthesis through weighting by the criteria and adding to capture order dominance. Otherwise, not only it may not yield an optimum order, but also could contradict the ranking obtained by hierarchic composition.

For those who may be concerned about the linearity of certain aspects of hierarchic order, we note that in the end the outcome of a decision is a multilinear form, a covariant tensor, that can have a highly nonlinear interpretation.

Open and Closed Structures - Rank Preservation and Reversal

What happens to the rank of alternatives if new alternatives are added or old ones deleted? This subject we discuss in detail in a separate paper included in these proceedings. We close by offering a diagram of four similar procedures for using the AHP to deal with single shot and sequential decisions. we refer to a decision problem as closed or open depending on whether all the alternatives are known in advance, or new ones are added as for example in student admissions, where decisions are made sequentially.

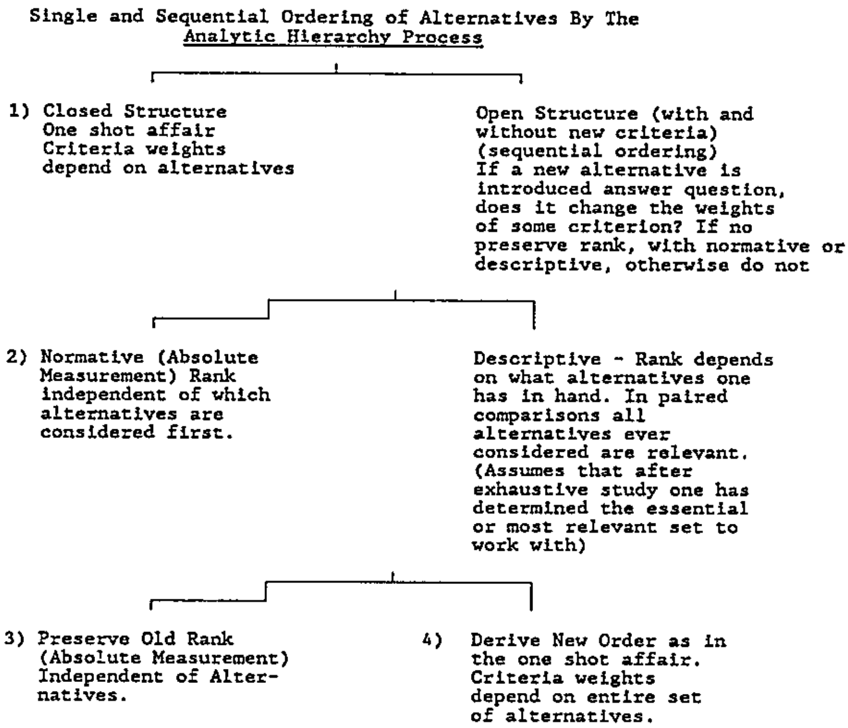

Figure 1

The normative approach has no way to include the effect of new or old alternatives on criteria weights. Each of these methods is needed for some problem. Rank can change if new alternatives introduce new criteria or change the weights of old criteria.

In descriptive absolute measurement a new alternative is compared with some other alternative, perhaps the highest ranking one with respect to a criterion and placed in its proper order. Any other alternative continues to be compared with the original ideal in the same manner.

References

1. Birkhoff, G., "Extensions of Jentzsch's Theorem", Trans. Amer. Math. Soc., 85, 219-227, 1957.

2. Kohlberg, E. and J.W. Pratt, "The Contraction Mapping Approach to the Perron-Frobenius Theory : Why Hilbert's Metric?", Mathematics of Operations Research, 7, 2, 198-210, 1982 .

3. Saaty, T.L., The Analytic Hierarchy Process, RWS Publications, 4922 Ellsworth Avenue, Pittsburgh, Pennsylvania, 15213, 1990.