ISAHP 1996, Vancouver, Canada, July 12-15, 1996

THE ANALYTIC HIERARCHY PROCESS AND UTILITY THEORY: RATIO SCALES AND INTERVAL SCALES

Rozann Saaty
Expert Choice, Inc.

Abstract: This paper is a brief summary of the successful workability of the Analytic Hierarchy Process (AHP) free of problems and paradoxes. As a theory of ratio scale measurement, the AHP has found multiple and diverse applications in priority setting, decision making, planning, resource allocation, forecasting and prediction, and in conflict resolution. Contrary to prevailing theory about impossibility, the AHP approach has been shown to lead to the possibility of constructing a reality capturing a group priority function which also captures the true preferences of the individuals and allows for different individual powers. Using the AHP, one can deduce Bayes' Theorem, considered important and basic in probability prediction particularly in medical diagnosis, but not without limitations and difficulties. Rank reversals found to happen frequently in practice are allowed for in the theory of the AHP, but not for example in Utility Theory. This paper is about the far reach of the AHP with an absence of paradoxes.

Introduction

The question is sometimes asked: What can one do with the new ratio scale approach to decision making of the Analytic Hierarchy Process (AHP) that can't be done better with classical Utility Theory? There are several areas in which the strengths of the AHP and its use of ratio scales have been demonstrated to work in an ideal way free of problems and paradoxes and fits well within the scientific method. We shall briefly discuss these areas. We will assume that we are writing for people knowledgeable about both the AHP and Utility Theory.

The Analytic Hierarchy Process (AHP) is a general theory of measurement. It is used to derive ratio scales from both discrete and continuous paired comparisons. These comparisons may be taken from actual measurements or from a fundamental scale which reflects the relative strength of preferences and feelings. The AHP has a special concern with departure from consistency, its measurement and on dependence within and between the groups of elements of its structure. It has found its widest applications in multicriteria decision making (Saaty, 1990), planning (Saaty and Kearns, 1991) in resource allocation (Saaty and Vargas, 1985), in forecasting and prediction (Saaty and Vargas, 1991), and in conflict resolution (Saaty and Alexander, 1989). In its general form, the AHP is a nonlinear framework for carrying out both deductive and inductive thinking without use of the syllogism by taking several factors into consideration simultaneously and allowing for dependence and for feedback, and making numerical trade-offs to arrive at a synthesis or conclusion.

For a long time, people have been concerned with the measurement of both physical and psychological events. By physical we mean the realm of what is fashionably known as the tangibles as it relates to some kind of objective reality outside the individual conducting the measurement. By contrast, the psychological is the realm of the intangibles as it relates to subjective ideas and beliefs of the individual, about himself or herself, and the world of experience. The question is whether there is a coherent theory that can deal with both these worlds of reality without compromising either. The AHP is a method that can be used to establish measures in both the physical and social domains.

In using the AHP to model a problem, one needs a hierarchic or a network structure to represent that problem and 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

(Saaty, 1994), from which ratio scales are derived in the form of principal eigenvectors, or eigenfunctions, as the case may be. These matrices, or kernels, are positive and reciprocal, e.g. $a_{ij} = 1/a_{ji}$. In particular, special efforts have been made to characterize these matrices. Because of the need for a variety of judgments, there has also been considerable work done to characterize the process of synthesizing diverse judgments.

Scales

There are several types of numerical scales that may be considered to rank criteria and alternatives in decision analysis. There are ordinal scales, invariant under strictly monotone increasing transformations; interval scales, invariant under positive linear transformations; ratio scales, invariant under positive similarity transformations; and absolute scales, invariant under the identity transformation. If they all lead to the same result it would not matter which is used and the distinction among scales would be superfluous. When there are multiple criteria, however, one cannot simply use any scale. It must be possible to represent judgments numerically in a meaningful and accurate way from which a ranking of the alternatives of a decision is produced. It must also be possible to combine or synthesize these rankings with respect to the different criteria, and not every scale allows the arithmetic operations needed to do the representation and synthesis meaningfully and accurately. Ordinal numbers, for example, are not serious contenders in this process. In addition, there are situations of interdependence among the alternatives that narrow the choice of scale further. We need to consider what numerical scales there are and whether arithmetic operations on them result in meaningful scales. Note that one cannot multiply numbers from an interval scale because the result is not an interval scale. Thus, $(ax_1+b)(ax_2+b) = a^2x^2+ab(x_1x_2)+b^2$ which does not have the precise form $ax+b$. One can take the average of interval scale readings but not their sum. Thus, $(ax_1+b)+(ax_2+b) = a(x_1+x_2)+2b$, which does not have the form $ax+b$. However, if we average by dividing by 2, we do get $a(x_1+x_2)/2 + b$ an interval scale value of the same form. Similarly, we can multiply interval scale readings by positive numbers whose sum is equal to one and add to get an interval scale result, a weighted average. For a ratio scale, we have $ax_1+ax_2 = a(x_1+x_2) = ax_3$ which belongs to the same ratio scale, and $ax_1bx_2 = abx_1x_2 = cx_1x_2 = cx_3$ which again belongs to a new ratio scale. However, ax_1+bx_2 does not define a ratio scale and, thus, we cannot add measurements from different ratio scales. These observations are the basis for what follows in this paper.

In multicriteria decisions based on interval or ratio scales, it is not reasonable to expect to choose a different scale for measuring the alternatives with respect to each criterion and yet to arrive at a unique overall decision. A way is required to represent the rankings with respect to each criterion on the same underlying ratio scale (priority theory, based on the dominance of preferences) or interval scale (utility theory). The latter is only valid for weighting the alternatives with criteria weights represented by a ratio scale. One cannot measure criteria or goals on an interval scale and then use them for weighting alternatives since one then obtains a product of two interval scales, a meaningless number. In any event, there is the problem of intangible criteria and how to construct their measurement. The next fundamental question is: How are numbers introduced into the analysis of decision making?

The Measurement of Tangibles and Intangibles

Judgment expressed numerically can be used to capture magnitude very closely. There are numerous examples to illustrate this point. These examples demonstrate the ability of the mind to correctly assess numerical magnitudes that characterize physical reality. Here is one of them.

Figure 1 shows five areas to which we can apply to the paired comparison process in a matrix and use the 1—9 scale to test the validity of the procedure. We can approximate the priorities in the matrix by assuming that it is consistent. We normalize each column and then take the average of the corresponding entries in the columns.

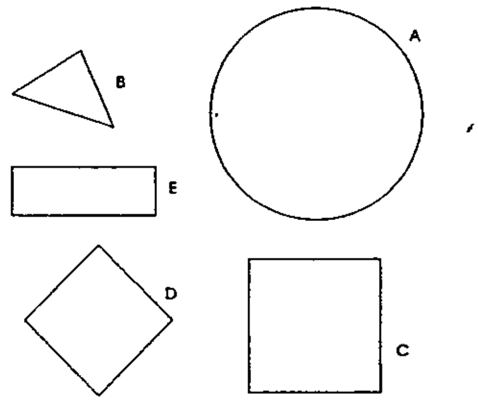

Figure 1: Five figures drawn with appropriate size of area

The object is to compare the areas in pairs to reproduce their relative weights. The actual relative values of these areas are A=0.47, B=0.05, C=0.24, D=0.14, and E=0.09 with which the answer may be compared. By comparing more than two alternatives in a decision problem, one is able to obtain better values for the derived scale because of redundancy in the comparisons, which helps improve the overall accuracy of the judgments.

The question is whether the mind can also estimate intangibles correctly along the same lines. The only way to validate such results would be to find out how well they conform with our expectations. In addition, we note that the mind synthesizes both tangible and intangible data to form an overall understanding of reality and must be faithful to its purpose by mixing all such information in the same way. This becomes obvious when we realize that the tangibles of today were the intangibles of yesterday. People have been coping with complexity successfully since the dawn of history before the arrival of modern science and its measurement of intangibles. Paired comparisons are an intrinsic biological capability to understand the world and not a convenient clever invention.

Three Paradoxes of Utility Theory

There are three kinds of paradoxes in utility theory. A major concern is that these paradoxes occur frequently and are of serious concern for the healthfulness of the theory. The first (typified by Allais' paradox) are examples contradicting the axiom of transitivity (axiom 4, which says that if an act belongs to a class of acts, and if another act is either preferred or indifferent or is simply indifferent to that act, then it also belongs to the class of acts). If A is preferred to B and B preferred to C, then A must be preferred to C. The second are examples contradicting the axiom of nonoptimal acts or alternatives (axiom 6, which says that adding new acts to a decision problem under uncertainty, each of which is weakly dominated by or is equivalent to some old act, has no effect on the optimality or non-optimality of an old act). An optimal alternative cannot be made non-optimal by introducing a new dominated (non-optimal) alternative. The third are examples contradicting the axiom of independence from irrelevant alternatives (axiom 7, which says that a non-optimal act cannot be made optimal by adding new acts to the problem). A non-optimal alternative cannot be made optimal by introducing a new alternative. Here is an example of each type.

1. Luce and Raiffa (1957, p. 22) point out that preferences must precede the numerical characterization of them. Then they say:

> If we add more gambles to the collection and try to assign utilities as we have done, it is clear that to be successful the subject's preferences will have to satisfy some consistency requirements. For example, if he prefers A to B, B to C, and a lottery which yields A with probability 2/3 and C with probability 1/3 to a lottery which yields A with probability 3/4 and C with probability 1/4, then we are in trouble. Or if he prefers A to B, B to C, and B to any lottery involving A and C as prizes so long as it is a bona fide gamble, i.e., $p \neq 1$, we are again in trouble.

Note that transitivity requires that the subject prefer A to C, and hence choose a lottery where the probability of A occuring relative to C occuring is the highest. But the subject does not do that, indicating a preference for C over A. To correct the situation, they propose imposing conditions (through axiomatization) on the way people should be thinking in rating alternatives and on the nature of the alternatives themselves.

2. Luce and Raiffa (1957, p. 288) give an example which violates axiom 6 that adding a dominated (less preferred) alternative cannot affect the rank of an old alternative. At a restaurant of unknown quality, a man who loves and can afford steak, when offered less expensive broiled salmon or more expensive steak, orders salmon rather than risking paying double the price of salmon for a steak of questionable quality. He is then quickly told, with an apology, that the restaurant also has fried snails and frog legs at a price comparable to that of the steak. The man shudders quietly at the thought of eating them, but then changes his order from salmon to steak. He reasons that this is a restaurant of high culinary discrimination and would serve a good steak. Thus, the presence of a dominated alternative (snails and frog legs, which he hates) can affect the rank of an old one contradicting axiom 6. In the AHP, by allowing a decision maker to consider separately a hierarchy of benefits, costs, opportunities, and risks, it is possible to perceive the phenomenon as natural because benefits and risks are regarded as two separate concerns. In this case, we assume that the decision is based only on benefits and risks. The new information about the availability of snails and frog legs will not alter the decision maker's preference with respect to benefits but will with respect to risks. The risks hierarchy would now show that the risk is small that the restaurant does not know to serve a good steak. By combining benefits and risks, rank reversal would occur.

3. A practical example which violates axiom 7, that the rank of a low-ranking alternative cannot be improved (or diminished against old acts) by adding new alternatives (whether dominated or not) is illustrated by the lady who goes hat shopping and finds two hats which she singles out and then selects A over B. When she visits another shop with many copies of A, she returns and buys B. (Abundance has decreased her preference for A.)

Group Decisions: The Possibility of Arrow's Conditions

When a group of individuals wants to make a collective choice, they would each express their own individual choices and one would then combine their choices into a group choice. One way to do this is to see what alternative has the most votes followed by the one with the next most votes and so on. But there are other ways of deriving a group choice from individual choices. It seems reasonable that no matter how we aggregate individual into group choice, if all the people prefer one alternative over another, then the group also should. This is called Pareto optimality. It is possible when there are more than two choices, the aggregated group order of choices may be intransitive, and thus it would be impossible to say which is the most preferred choice. This is known as the Condorcet paradox. Borda's answer to this paradox is for each individual to number his choices from *1* to *n* and the group choice is made by summing the numbers assigned by the individuals to each alternative. In case of a tie, the alternatives involved are arbitrarily ordered among themselves. An objection to Borda's scheme is why use equidistant numbers when one alternative may be much preferred by one person but a little by another. Kenneth Arrow proved that in general it is impossible to arrive at a rational group choice from individual *ordinal* preferences. Rationality implies that four conditions should be satisfied:

1. Non-dictatorship: no single individual member of the group determines the group order.
2. Decisiveness: the aggregation procedure must generally produce a group order.
3. Pareto optimality: if every individual prefers A to B, then so does the group.
4. Independence of irrelevant alternatives: the group choice between two alternatives must be based on the individual preferences only between that pair of alternatives.

Arrow's impossibility is made possible by using cardinal numbers as in the AHP, as shown by Kirti Peniwati (1996) in her Ph.D. dissertation. The foregoing conditions are satisfied in the AHP if one aggregates the separate orderings by the individuals using the geometric mean. However, in the AHP one can also aggregate judgments to obtain a group choice. If these judgments are inconsistent, the individuals need to work on improving inconsistency. The resulting group order may be different from the previous order, but also satisfies the above four conditions.

AHP & Bayes Theorem

A popular statistical approach in medical diagnosis is Bayes Theorem. It provides a paradigm for updating diagnostic information encoded in the form of probabilities. It involves the assumption that decisions involving uncertainty can only be made with the aid of information about the environment in which the decision is made. Bayes theory updates information by using Bayes theorem, a statement in conditional probabilities relating causes (states of nature) to outcomes. Outcomes are results of experiments used to uncover the causes. Bayes theory revises initial or prior probabilities of causes known from a large sample of a population, into posterior probabilities by using the outcome of an experiment or test with a certain probability of success performed on a particular individual. Prior probabilities are obtained either subjectively or empirically by sampling the frequency of occurrence of a cause in a population. Posterior probabilities are those based on the prior probabilities, and on both the outcome of the experiment and on the observed reliability of that experiment.

Judgments are needed in medical diagnosis to determine which test to perform given certain symptoms. Information on symptoms and combinations of symptoms (Saaty, 1994) is not well known and easily recognized for many diseases. To apply Bayes theorem in this context in a meaningful way, one needs statistical data from experiments or expert knowledge relating the symptoms to the diseases. Even when the number of symptoms is small, the required number of experiments to generate adequate statistical data can be unmanageably large. This way of addressing the diagnostic problem with combinatorics reminds one of computer chess programs which search large portions of the space of possible chess moves and assign to each of the moves values according to some expert judgment.

There is a similar need in diagnosis for a model that incorporates both statistical data and expert judgment. When statistical data are present but no expert judgment is available, one property of this model should be to reproduce results obtained through time-honored procedures such as Bayes theorem. When expert judgment is present, it should be possible to combine judgment with statistical data to identify the disease which best describes the observed symptoms.

Uncertainty in statistics is represented with probabilities, measured on an absolute scale with zero origin, and an absolute scale can be used to construct a ratio scale (invariant under positive similarity transformations). In addition, it is known in decision theory that judgments can be represented in relative terms with ratio scales. The desired model should be a generalization that can account for the workings of both the statistical model and the judgment model. This implies that the model must be based on a unifying theory of measurement that can equally deal with and combine statistical and judgmental data.

In principle, the ratio scales resulting from comparisons of statistical data could be taken as fine structured judgments applied to the underlying information. Thus, statistics and judgments can be combined through a common ratio scale. That is what the judgment and data-based multicriteria decision theory, the Analytic Hierarchy Process (AHP) does. Here we are interested in the AHP framework that deals with dependence among the elements or clusters of a decision structure to combine statistical and judgmental evidence.

Rank Reversal and Utility Rethinking

Utility theory does not allow for rank reversal which has been observed to occur both in theory and in practice by many scholars in utility theory itself. For those decisions in which rank should reverse, utility theory cannot give the right answer. Some scholars in utility theory have talked about rethinking the decision problem; but what can rethinking the problem mean? Assume that in a particular decision no new criteria are introduced nor their importance changed by the new alternatives. What would rethinking the problem do? One cannot disconnect from a given formulation for technical convenience. If the problem is reformulated differently in order to obtain the desired rank, one cannot consider the process of rethinking as really solving the problem of rank reversal. It would appear that the technical problem of allowing rank to change has not been solved in this way. There are examples where the same structure with the same number of criteria and alternatives, but with different names, and with the same set of judgments whereby in one decision problem one needs to preserve rank and in the other one needs to allow rank to reverse. The familiar hats problem could be reformulated with the names of the criteria appropriately changed to apply to computers, and the same judgments are used to rank the criteria and the computers. For the hats, the buyer may feel that the presence of many copies of the preferred hat would change her preference to the second preferred hat, whereas for the computers the presence of copies would not. Thus the decision is up to the buyer as to whether the alternatives are ranked according to their performance on a standard, or whether the best in the group is desired. Note that manyness is not a property of any single thing and cannot be included as a criterion because it implies dependence among the alternatives. Some utility theorists have suggested that they would rethink the problem. How would rethinking help solve such a problem? One cannot play games with a theory. Either it works, or it does not. In the AHP, rank preservation and reversal are a part of the decision to be made and there is a procedure for preserving rank and another for allowing rank to reverse.

Conclusion

The use of ratio scales leads to a theory for decision making and generalizations of that theory that neither require more complex formalization nor lead to paradoxes. The Analytic Hierarchy Process and its generalization, the Analytic Network Process are the realization of the ratio scale approach in decision making. Utility theory has yet to solve its paradoxes and allow for dependence and feedback in complex decisions.

References

Saaty, Thomas L. (1990), The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation. Pittsburgh, PA: RWS Publications.

Saaty, Thomas L., and Kevin P. Kearns (1991), Analytical Planning: The Organization of Systems. Pittsburgh, PA: RWS Publications.

Saaty, Thomas L, and Luis G. Vargas (1991), The Logic of Priorities: Applications of the Analytic Hierarchy Process in Business, Energy, Health & Transportation. Pittsburgh, PA: RWS Publications,

Saaty, Thomas L., and Luis G. Vargas (1991), Prediction, Projection and Forecasting. Norwell, MA: Kluwer Academic Publishers.

Saaty, Thomas L., and Joyce M. Alexander (1989), Conflict Resolution: The Analytic Hierarchy Approach. New York, NY: Praeger Publishers,

Luce, R. Duncan, and Howard Raiffa (1957), Games and Decisions. New York: Wiley.

Peniwati, K (1996) "The Possibility Theorem for Group Decision Making: The Analytic Hierarchy Process", Ph.D Dissertation, Katz Graduate School of Business, University of Pittsburgh.

Saaty, Thomas L. (1994), Fundamentals of Decision Making and Priority Theory with The Analytic Hierarchy Process, Pittsburgh, PA: RWS Publications.