ДІТ
ПРОБЛЕМИ ПРИЙНЯТТЯ РІШЕНЬ І УПРАВЛІННЯ В ЕКОНОМІЧНИХ, ТЕХНІЧНИХ, ЕКОЛОГІЧНИХ І СОЦІАЛЬНИХ СИСТЕМАХ

УДК 519.5

THEORY OF THE ANALYTIC HIERARCHY PROCESS. PART 2.1

THOMAS L. SAATY

Highlights are given of the decision making theory the Analytic Hierarchy Process (AHP) and its generalization to dependence and feedback, the Analytic Network Process (ANP) both of which deal with the measurement of tangible and intangible criteria in relative terms by using paired comparisons. The fundamental scale of absolute numbers for representing judgments is introduced and the principal right eigenvector is shown to be the necessary vector of priorities derived from the possibly inconsistent matrix of comparisons of homogeneous elements. A method of synthesis of priorities is proposed. Rank preservation and reversal are discussed. When independent, the alternatives can be rated one at a time with respect to the criteria using intensities. Negative priorities are also introduced. This paper will be followed soon by two other papers.

1. INTRODUCTION

Not long ago, people believed that the human mind is a completely unreliable instrument to perform measurement and that the only meaningful measurement is obtained on a physical measurement scale like the meter and the kilogram invented by some clever person. They did not think how the measurements came to have meaning for people and that this meaning depends on people's purpose each time they obtain a reading on that scale. In the winter, ice may be a source of discomfort but an ice drink in the summer can be a refreshing source of comfort. A number has no meaning except that assigned it by someone's mind. We may all agree on the number we read on a physical scale, but not on what exactly that number means to us in practical terms. We tend to parrot abstractions that define a number but not how we use that number, which is ultimately more important for our survival. Thus it is our subjective values that are essential for interpreting the readings obtained through measurement. This interpretation depends on what one has in mind at the time and different people may interpret the same reading differently for the same situation depending on their goal. The reading may be called objective but the interpretation and use are both subjective. In this sense subjectivity is important because without it objectivity has no intrinsic meaning. If the mind of an expert can produce measurement close to what we obtain through measuring instruments, then it has greater power than the instruments in dealing with complexity because it can also measure things for which we have no instrument. What we have to do is examine the possibility and validity of this assumption as critically as we can. It turns out that when we have knowledge and experi-

© *Thomas L. Saaty,* 2003

ence, our minds are very good measuring instruments. This does not mean that we should discard what we use in science, but rather use it to support and strengthen what we are born to do with our minds.

The basic problem is that we need to measure intangibles of which there is a near infinite number and we can only do it by making comparisons in relative terms. Even if everything were measurable, we would still need to compare the different types of measurement and how important they are to us in some decision in order to make tradeoffs among them and reach a final answer. If we use tangibles and their measurements we would need to normalize them to a common frame of reference to reduce them to relative numbers and then weight and combine them with intangibles. Combining normalized priorities of measurable quantities with normalized relative measurement of non-measurable things needs ratio scales because we can weight and add the outcomes.

The Analytic Hierarchy Process (AHP) is a theory of relative measurement on absolute scales of both tangible and intangible criteria based on the judgment of knowledgeable and expert people. How to measure intangibles is the main concern of the mathematics of the AHP. In the end we must fit our entire world experience into our system of priorities if we are going to understand it. The AHP is based on four axioms: (1) reciprocal judgments, (2) homogeneous elements, (3) hierarchic or feedback dependent structure, and (4) rank order expectations. The synthesis of the AHP combines multidimensional scales of measurement into a single "unidimensional" scale of priorities. Decisions are determined by a single number for the best outcome or by a vector of priorities that gives an ordering of the different possible outcomes. We can also combine our judgments or our final choices obtained from a group when we wish to cooperate to agree on a single outcome.

The AHP has been mostly applied to multi-objective, multi-criteria and multiparty decisions because decision making needs this diversity. To make tradeoffs among the many objectives and many criteria, the judgments that are usually made in qualitative terms are expressed numerically. To do this, rather than simply assigning a score out of a person's memory that appears reasonable, one must make reciprocal pairwise comparisons in a carefully designed scientific way. In paired comparisons the smaller or lesser element is used as the unit, and the larger or greater element is estimated as a multiple of that unit with respect to the common property or criterion for which the comparisons are made. In this sense measurement with judgments is made more scientifically than by assigning numbers more or less arbitrarily. Because human beings are limited in size and the firings of their neurons are limited in intensity, it is clear that there is a limit on their ability to compare the very small with the very large. It is precisely for this reason that pairwise comparisons are made on elements or alternatives that are close or homogeneous and the more separated they are, the more need there is to put them in different groups and link these groups with a common element from one group to an adjacent group of slightly greater or slightly smaller elements.

From all the paired comparisons, one derives a scale of relative values for the priorities. Because of inconsistency among the judgments and more importantly because of the need for the invariance of priorities, it is mathematically necessary to derive the priorities in the form of the principal eigenvector of the matrix of paired comparisons.

We learn from making paired comparisons in the AHP that if A is 5 times larger than B and B is 3 times larger than C, then A is 15 times larger than C and A dominates C 15 times. That is different from A having 5 dollars more than B and B having 3 dollars more than C implies that A has 8 dollars more than C. Defining intensity along the arcs of a graph and raising the matrix to powers measures the first kind of dominance precisely and never the second. It has definite meaning and as we shall see below, in the limit is measured uniquely by the principal eigenvector. The use of ratios to represent dominance leads to this idea of dominance and can be verified by taking a matrix whose entries are written as ratios, raising it to powers and checking that the resulting coefficients give the desired dominance from any point to any other. There is a useful connection between what we do with dominance priorities in the AHP and what is done with transition probabilities both of which use matrix algebra to find their answers. Transitions between states are multiplied and added. To compose the priorities for the alternatives of a decision with respect to different criteria, it is also necessary that the priorities of the alternatives with respect to each criterion be multiplied by the priority of that criterion and then added over all the criteria.

The rank of the alternatives can be affected by how many alternatives there are and also by the quality or measurement of those alternatives. Both of these are not intrinsic properties of any alternative, but can affect the importance of that alternative. It may at first seem convenient to assume that alternatives are independent of each other; but that is not always natural. For example as we increase the number of copies of an alternative, it often loses (and sometimes increases) its importance. For example, if gold were an alternative and gold is important, and if we increase the amount of gold available until it is near infinite, gold gradually loses its importance. No new criterion is added and no judgment is changed. Relative measurement takes quantity into consideration. We often need to consider this kind of dependence known as structural dependence. When we add more alternatives, the ranks among old ones may change and what was preferred to another now because of the presence of new ones may no longer be preferred to the other. Another example is that of a company that sells cars A and B. Car B is better than car A but it costs more to make. It is more desirable all around for people to buy car B but they buy A because it is cheaper. The company advertises that it is going to make car C that is similar to B but much more expensive. People are now observed more and more to buy car B. The company never makes car C. This is a real life example from marketing. However, in some decision problems we may want to treat by fiat the alternatives of a decision as completely independent both in property and in number and quality and want to preserve the ranks of existing alternatives when new ones are added or old ones deleted. The AHP allows for both these possibilities. Actually, change in rank in the presence of relevant alternatives is a fact of our world. It is also a fact that when the number of irrelevant alternatives is very large, they can cause rank to change. Viruses are irrelevant in most decisions but they can eventually cause the death of all decision makers and make mockery of the decisions they thought were so important. In essence reality is much more interdependent than we have allowed for in our limited ways of thinking. Admittedly there are times when we wish to preserve rank no matter what the situation may be. We need to allow for both in our decision theories and not take the simple way out as most theories do at present.

Our story must begin with how paired comparisons are made, what scales of numbers to use and what kind of scale to derive from them and how. The AHP has three parts: what a person has to do; what the method has to do; and what the decision process has to do to deliver "right" answers. The decision maker needs to:

1. Know his problem well and know how to structure it;

2. Put the more important criteria up in a hierarchy, the subcriteria below that and the alternatives below that; or in a network make clusters of elements and then connect these clusters and elements in them according to influence as in a hierarchy with the arrow pointing in the opposite direction to how one determines dominance of influence;

3. Provide judgments in answering two kinds of questions that we know in reality that everyone has a talent to do:

   a. With respect to a property or criterion which of two elements is more **dominant**, and
   b. With respect to a control criterion and a criterion in a subnet, which of two elements contributes more to or influences the control criterion more through that criterion in the subnet? This is a conditional kind of question and very important to keep in mind.

Then there are three things to know about the AHP method of setting priorities.

1. We can derive the fundamental integer scale 1–9 of the AHP that we use to express judgments on pairs of elements from stimulus-response theory, and then validate it with numerous examples in practice whose measurements are known. We can use the measurements themselves instead of judgments if we want. We can connect measurement from the small to the very large through clustering to extend the 1–9 scale indefinitely;

2. Use a pairwise positive reciprocal comparisons matrix and derive the principal right eigenvector as the vector of priorities when the matrix is consistent. We use perturbation arguments to show that we must still solve the principal eigenvalue problem when it is reasonably but not greatly inconsistent. We also show that we need the principal eigenvector to capture transitivity by raising the matrix to powers. Finally because a hierarchy is a special case of a supermatrix raised to powers to capture the transitivity of influences, raising it to powers involves the arithmetic operations of multiplying and adding. In the limit we obtain the principal eigenvector as a necessary condition to preserve the **invariance** of the priorities. This means that if we use these priorities to weight the columns of the matrix to derive priorities from it, we obtain the original priorities back;

3. In both a hierarchic and a network structure of a decision, the priorities can be arranged as parts of its columns in a **supermatrix** [1, 2] whose clusters must be weighted according to their influence on each of them; the weights thus derived are used to weight the blocks of the supermatrix corresponding to the components to make it stochastic so it can be raised to powers to capture the interactions and obtain the priorities in the limit. There are two possibilities here depending on the irreducibility/reducibility of the matrix. The first yields a unique outcome while the second involves a cycle of limiting outcomes that are combined into an average outcome.

Finally the process of decision-making requires us to analyze a decision according to benefits (**B**), the good things that would result from taking the decision; opportunities (**O**), the potential good things that can result in the future from taking the decision; costs (**C**), the pains and disappointments that would result from taking the decision; and risks (**R**), the potential pains and disappointments that can result from taking the decision. We then create control criteria and subcriteria or even a network of criteria under each and develop a subnet and its connection for each control criterion. We then need to determine the best outcome for each control criterion and combine the alternatives in what is known as the ideal form for all the control criteria under each of **BOCR**. Then we note the best alternative under **B** and use it to think of benefits and then the best under **O**, which may be different than the one under **C**, and use it to think of costs and so on. Finally we must rate these four with respect to strategic criteria using the absolute mode of the AHP to obtain priority ratings for **B**, **O**, **C**, and **R**. We then normalize and use these weights to combine the four vectors of outcomes under **BOCR** to obtain an overall answer. We can form the ratio **BO**/**CR** which does not need the **BOCR** ratings to obtain marginal overall outcomes. Alternatively:

1) we can use the ratings to weight and subtract the costs and risks from the sum of the weighted benefits and opportunities, or

2) we can subtract the costs and the risks from one and weight and add everything, or

3) we can take the reciprocals of the costs and the risks, normalize the result and weight and add them to the weighted benefits and costs.

2. PAIRWISE COMPARISONS

Assume that one is given *n* stones, $A_1,...,A_n$, with known weights $w_1,...,w_n$, respectively, and suppose that a matrix of pairwise ratios is formed whose rows give the ratios of the weights of each stone with respect to all others. Thus one has the equation:

$$Aw = \begin{array}{c} \\ A_1 \\ \vdots \\ A_n \end{array} \begin{array}{c} A_1 \quad \cdots \quad A_n \\ \begin{bmatrix} \frac{w_1}{w_1} & \cdots & \frac{w_1}{w_n} \\ \vdots & \cdots & \vdots \\ \frac{w_n}{w_1} & \cdots & \frac{w_n}{w_n} \end{bmatrix} \end{array} \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix} = n \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix} = n\,w,$$

where A has been multiplied on the right by the vector of weights w. The result of this multiplication is $n\,w$. To recover the scale from the matrix of ratios, one must solve the problem $A = n\,w$ or $(A - n\mathbf{I})w = 0$ (where $\mathbf{I}$ is identify matrix). This is a system of homogeneous linear equations. It has a nontrivial solution if and only if the determinant of $A - n\mathbf{I}$ vanishes, that is, n is an eigenvalue of A. Now A has unit rank since every row is a constant multiple of the first row. As a result, all its eigenvalues except one are zero. The sum of the eigenvalues of a matrix is equal to its trace, the sum of its diagonal elements, and in this case the trace of A is equal to n. Thus n is an eigenvalue of A, and one has a nontrivial solution. The solution consists of positive entries and is unique to within a multiplicative constant.

To make w unique, we can normalize its entries by dividing by their sum. Thus, given the comparison matrix, we can recover the scale. In this case, the solution is any column of A normalized. Notice that in A the reciprocal property $a_{ji} = 1/a_{ij}$ holds; thus, also $a_{ii} = 1$. Another property of A is that it is consistent: its entries satisfy the condition $a_{jk} = a_{ik}/a_{ij}$. The entire matrix can be constructed from a set of n elements that form a chain across the rows and columns of A.

In the general case, the precise value of w_i/w_j cannot be given, but instead only an estimate of it as a judgment. For the moment, consider an estimate of these values by an expert whose judgments are small perturbations of the coefficients w_i/w_j. This implies small perturbations of the eigenvalues.

Let us for generality call $A_1, A_2, ..., A_n$ stimuli instead of stones. The quantified judgments on pairs of stimuli A_i, A_j, are represented by an $n \times n$ matrix $A' = (a_{ij}),\ \ i, j = 1, 2, ..., n$. The entries a_{ij} are defined by the following entry rules.

Rule 1. If $a_{ij} = a$, then $a_{ij} = 1/a,\ \ a \neq 0$.

Rule 2. If A_i is judged to be of equal relative intensity to A_j then $a_{ij} = 1$, $a_{ji} = 1$; in particular, $a_{ii} = 1$ for all i.

Thus the matrix A' has the form

$$A' = \begin{bmatrix} 1 & a_{12} & \cdots & a_{1n} \\ 1/a_{12} & 1 & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ 1/a_{1n} & 1/a_{2n} & \cdots & 1 \end{bmatrix}.$$

Having recorded the quantified judgments on pairs of stimuli (A_i, A_j) as numerical entries a_{ij} in the matrix A', the problem now is to assign to the n stimuli $A_1, A_2, ..., A_n$ a set of numerical weights $w_1, ..., w_n$ that would "reflect the recorded judgments." In order to do that, the vaguely formulated problem must first be transformed into a precise mathematical one. This essential, and apparently harmless, step is the most crucial one in any problem that requires the representation of a real-life situation in terms of an abstract mathematical structure. It is particularly crucial in the present problem where the representation involves a number of transitions that are not immediately discernible. It appears, therefore, desirable in the present problem to identify the major steps in the process of representation and to make each step as explicit as possible to enable the potential user to form his own judgment as to the *meaning and value* of the method in relation to *his* problem and *his* goal.

Why we must solve the principal eigenvalue problem in general has a simple justification based on the idea of dominance among the elements represented by the coefficients of the matrix. Dominance between two elements is obtained as the normalized sum of paths of all lengths between them. All the paths of a given length k are obtained by raising the matrix to the power k. According to Cesaro summability, the limit of the Cesaro sum $\lim_{N\to\infty} 1/N \sum_{k=0}^{N} A^k$ that represents the average of all order dominance up to N, is the same as the limit of the sequence of the powers of the matrix i.e. $\lim_{k\to\infty} A^k$. Now we know from Perron theory that this sequence converges to a matrix all whose columns are identical and are proportional to the principal right eigenvector of A.

Without the theory of Perron, the proof is somewhat more elaborate, but more interesting because it is related to the amount of inconsistency one allows.

The following theorem assures us that a sufficiently small perturbation $a_{ij} = \frac{w_i}{w_j}\varepsilon_{ij}$, $\varepsilon_{ij} > 0$, of a consistent matrix (which we know has a simple principal eigenvalue *n*) gives rise to a simple eigenvalue problem for an inconsistent matrix but does not guarantee that it is the principal eigenvalue.

Theorem *(Existence)*. If λ is a simple eigenvalue of A, then for small $\varepsilon > 0$, there is an eigenvalue $\lambda(\varepsilon)$ of $A(\varepsilon)$ with power series expansion in ε:

$$\lambda(\varepsilon) = \lambda + \varepsilon\lambda^{(1)} + \varepsilon^2 \lambda^{(2)} + \ldots$$

and corresponding right and left eigenvectors $w(\varepsilon)$ and $v(\varepsilon)$ such that

$$w(\varepsilon) = w + \varepsilon w^{(1)} + \varepsilon^2 w^{(2)} + \ldots$$

$$v(\varepsilon) = v + \varepsilon \lambda^{(1)} + \varepsilon^2 v^{(2)} + \ldots$$

Our general problem takes the form:

$$A'w = \begin{bmatrix} 1 & a_{12} & \cdots & a_{1n} \\ 1/a_{12} & 1 & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ 1/a_{1n} & 1/a_{2n} & \cdots & 1 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} = cw.$$

We know that the homogeneous system of linear equations $A'w = cw$ has a solution w if c is an eigenvalue of A'. We need to show that the priorities we seek should be given by solving the **principal** eigenvalue problem of A' and not by some other method. For example, the method of least squares (LSM) determines a vector by minimizing the Frobenius norm of the difference between A and a positive rank one reciprocal matrix $\lfloor y_i / y_j \rfloor$:

$$\min_{y>0} \sum_{i,j=1}^{n} (a_{ij} - y_i / y_j)^2 \ .$$

The method of logarithmic least squares (LLSM) determines a vector by minimizing the Frobenius norm of $\lfloor \log (a_{ij} x_j / x_i) \rfloor$:

$$\min_{x>0} \sum_{i,j=1}^{n} [\log a_{ij} - \log(x_i / x_j)]^2 .$$

Continuing with our quest for the principal right eigenvector, we need to take a small diversion to discuss the principle of invariance of priorities that provide a second necessary condition for deriving priorities.

The concept of invariance is one of the most central in mathematics. An invariant is a mapping of a set of elements with an equivalence relation, into another set of elements that is constant or unchanged with respect to the equivalence relation of the original set. Actually one speaks of an invariant of an element of the original set with respect to a mapping of that entire set. It should be clear that no matter what method we use to derive the weights w_i, we need to get them back as proportional to the expression $\sum_{j=1}^{n} a_{ij} w_j \quad i = 1,...n$, that is, we must solve $\sum_{j=1}^{n} a_{ij} w_j = c w_i$, $i = 1,...n$. Otherwise $\sum_{j=1}^{n} a_{ij} w_j$, $i = 1,...,n$ would yield another set of different weights and they in turn can be used to form new expressions $\sum_{j=1}^{n} a_{ij} w_j$, $i = 1,...,n$, and so on ad infinitum violating the need to have priorities that are invariant. Unless we solve the principal eigenvalue problem, our effort to derive priorities becomes meaningless.

We now show that the perturbed eigenvalue from the consistent case is the principal eigenvalue of A'. Our argument involves both left and right eigenvectors of A'. Two vectors $x = (x_1,...,x_n)$, $y = (y_1,...,y_n)$ are orthogonal if their scalar product $x_1 y_1 + ... + x_n y_n$ is equal to zero. It is known that any left eigenvector of a matrix corresponding to an eigenvalue is orthogonal to any right eigenvector corresponding to a different eigenvalue. This property is known as bi-orthogonality.

Theorem: For a given positive matrix A, the only positive vector w and only positive constant c that satisfy $Aw = cw$, is a vector w that is a positive multiple of the principal eigenvector of A, and the only such c is the principal eigenvalue of A

Proof: We know that the right principal eigenvector and the principal eigenvalue satisfy our requirements. We also know that the algebraic multiplicity of the principal eigenvalue is one, and that there is a positive left eigenvector of A (call it z) corresponding to the principal eigenvalue. Suppose there is a positive vector y and a (necessarily positive) scalar d such that $Ay = dy$. If d and c are not equal, then by bi-orthogonality y is orthogonal to z, which is impossible since both vectors are positive. If d and c are equal, then y and w re dependent since c has algebraic multiplicity one, and y is a positive multiple of w. This completes the proof.

Thus we see that both requirments of dominance and invariance lead us to the principal right eigenvector. The problem now is how good is the estimate of w. Notice that if w is obtained by solving this problem, the matrix whose entries are w_i / w_j is a consistent matrix. It is a consistent estimate of the matrix A'. The matrix A' itself need not be consistent. In fact, the entries of A' need not even be transitive; that is, A_1 may be preferred to A_2 and A_2 to A_3 but A_3 may be preferred to A_1. What we would like is a measure of the error due to inconsistency. It turns out that A' is consistent if and only if $\lambda_{\max} = n$ and that we always have $\lambda_{\max} \geq n$.

Remark. Solving the principal eigenvalue problem to obtain priorities is equivalent to the following two optimization problems: Find $w_i > 0$, $i = 1, \ldots, n$ which

1) maximize $\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}\frac{w_j}{w_i}$ or, in the simpler linear optimization setting;

2) maximize $\sum_{j=1}^{n} w_j \sum_{j=1}^{n} a_{ij}$, obtained by multiplying the sum of each

column *j* by its corresponding w_j and summing over *j*, subject to $\sum_{i=1}^{n} w_i = 1$.

3. STIMULUS RESPONSE AND THE FUNDAMENTAL SCALE

To be able to perceive and sense objects in the environment our brains miniaturize them within our system of neurons so that we have a proportional relationship between what we perceive and what is out there. Without proportionality we cannot coordinate our thinking with our actions with the accuracy needed to control the environment. Proportionality with respect to a single stimulus requires that our response to a proportionately amplified or attenuated stimulus we receive from a source should be proportional to what our response would be to the original value of that stimulus. If $w(s)$ is our response to a stimulus of magnitude s, then the foregoing gives rise to the functional equation $w(as) = bw(s)$. (This equation can also be obtained as the necessary condition for solving the Fredholm equation of the second kind

$$\int_a^b K(s,t)\, w(t)\, dt = \lambda_{\max}\, w(s)$$

obtained as the continuous generalization of the discrete formulation $Aw = \lambda_{\max}\, w$ for obtaining priorities where instead of the positive reciprocal matrix A in the principal eigenvalue problem, we have a positive kernel, $K(s,t) > 0$, with $K(s,t)\, K(s,t) = 1$ that is also consistent i.e. $K(s,t)\, K(t,u) = K(s,u)$, for all s, t and u. The solution of this functional equation in the real domain is given by

$$w(s) = C\,e^{\log b \frac{\log s}{\log a}}\, P\left(\frac{\log s}{\log a}\right),$$

where P is a periodic function of period 1 and $P(0)=1$. One of the simplest such examples with $u = \log s / \log a$ is $P(u) = \cos(u/2B)$ for which $P(0)=1$.

The logarithmic law of response to stimuli can be obtained as a first order approximation to this solution through series expansions of the exponential and of the cosine functions as:

$$v(u) = C_1 e^{-\beta u} P(u) \approx C_2 \log s + C_3 ,$$

where $\log ab = -\beta,\ \beta > 0$. The expression on the right is known as the Weber-Fechner law of logarithmic response $M = a \log s + b,\ a \neq 0$ to a stimulus of magnitude *s*. This law was empirically established and tested in 1860 by Gustav Theodor Fechner who used a law formulated by Ernest Heinrich Weber regarding discrimination between two nearby values of a stimulus. We have now shown that it can be derived that Fechner's version can be derived by starting with a functional equation for stimulus response.

The integer-valued scale of response used in making paired comparison judgments can be derived from the logarithmic response function as follows. For a given value of the stimulus, the magnitude of response remains the same until the value of the stimulus is increased sufficiently large in proportion to the value of the stimulus, thus preserving the proportionality of relative increase in stimulus for it to be detectable for a new response. This suggests the idea of just noticeable differences (jnd), well known in psychology. Thus, starting with a stimulus s_0 successive magnitudes of the new stimuli take the form:

$$s_1 = s_0 + \Delta s_0 = s_0 + \frac{\Delta s_0}{s_0} s_0 = s_0\,(1+r),$$

$$s_2 = s_1 + \Delta s_1 = s_1\,(1+r) = s_0\,(1+r)^2 s_0 \alpha^2 ,$$

$$s_n = s_{n-1}\alpha = s_0 \alpha^n \quad (n = 0,1,2,...).$$

We consider the responses to these stimuli to be measured on a ratio scale $(b=0)$. A typical response has the form $M_i = a \log \alpha^i,\ \ i = 1,...,n$, or one after another they have the form:

$$M_1 = a \log \alpha,\ M_2 = 2a \log \alpha, ...,\ M_n = n a \log \alpha .$$

We take the ratios $M_i / M_1,\ \ i = 1,...,n$ of these responses in which the first is the smallest and serves as the unit of comparison, thus obtaining the *integer* values $1,2,...,n$ of the fundamental scale of the AHP. It appears that numbers are intrinsic to our ability to make comparisons, and that they were not an invention by our primitive ancestors. We must be grateful to them for the discovery of the symbolism. In a less mathematical vein, we note that we are able to distinguish ordinally between high, medium and low at one level and for each of them in a second level below that also distinguish between high, medium and low giving us nine different categories. We assign the value one to (low, low) which is the smallest and the value nine to (high, high) which is the highest, thus covering the spectrum of possibilities between two levels, and giving the value nine for the top of the paired comparisons scale as compared with the lowest value on the scale. In fact we show later that because of increase in inconsistency when we compare more than about 7 elements, we don't need to keep in mind more than 7 ± 2 elements. This was first conjectured by the psychologist George Miller in the 1950's and explained in the AHP in the 1970's [4]. Finally, we note that the scale just derived is attached to the importance we assign to judgments. If we have an exact measurement such as 2.375 and want to use it as it is for our judgment without attaching significance to it, we can use its entire value without approximation.

A person may not be schooled in the use of numbers but still have feelings, judgments and understanding that enable him to make accurate comparisons (equal, moderate, strong, very strong and extreme and compromises between these intensities). Such judgments can be applied successfully to compare stimuli that are not too disparate but homogeneous in magnitude. By homogeneous we mean that they fall within specified bounds. The foregoing may be summarized as in Table 1 to represent the fundamental scale for paired comparisons.

Table 1. The Fundamental Scale of Absolute Numbers

Intensity of Importance	Definition	Explanation
1	**Equal Importance**	**Two activities contribute equally to the objective**
2	Weak	
3	**Moderate importance**	**Experience and judgment slightly favor one activity over another**
4	Moderate plus	
5	**Strong importance**	**Experience and judgment strongly favor one activity over another**
6	Strong plus	
7	**Very strong or demonstrated importance**	**An activity is favored very strongly over another; its dominance demonstrated in practice**
8	Very, very strong	
9	**Extreme importance**	**The evidence favoring one activity over another is of the highest possible order of affirmation**
Reciprocals of above	**If activity *i* has one of the above nonzero numbers assigned to it when compared with activity *j*, then *j* has the reciprocal value when compared with *i***	**A reasonable assumption**
Rationals	**Ratios arising from the scale**	**If consistency were to be forced by obtaining *n* numerical values to span the matrix**

In a judgment matrix A instead of assigning two numbers w_i and w_j (that generally we do not know), as one does with tangibles, and forming the ratio w_i / w_j we assign a single number drawn from the fundamental scale of absolute numbers shown in Table 1 above to represent the ratio $(w_i / w_j)/1$. It is a nearest integer approximation to the ratio w_i / w_j. The ratio of two numbers from a ratio scale (invariant under multiplication by a positive constant) is an absolute number (invariant under the identity transformation). The derived scale will reveal what w_i and w_j are. This is a central fact about the relative measurement approach. It needs a fundamental scale to express numerically the relative dominance relationship.

Remark: The reciprocal property plays an important role in combining the judgments of several individuals to obtain a judgment for a group. Judgments must be combined so that the reciprocal of the synthesized judgments must be equal to the syntheses of the reciprocals of these judgments. It has been proved that the geometric mean is the unique way to do that. If the individuals are experts, they my not wish to combine their judgments but only their final outcome from a hierarchy. In that case one takes the geometric mean of the final outcomes. If the individuals have different priorities of importance their judgments (final outcomes) are raised to the power of their priorities and then the geometric mean is formed [3].

Validation Examples

Fig. 1 shows five areas to which we can apply the paired comparison process in a matrix and use the 1 to 9 scale to test the validity of the procedure. The object is

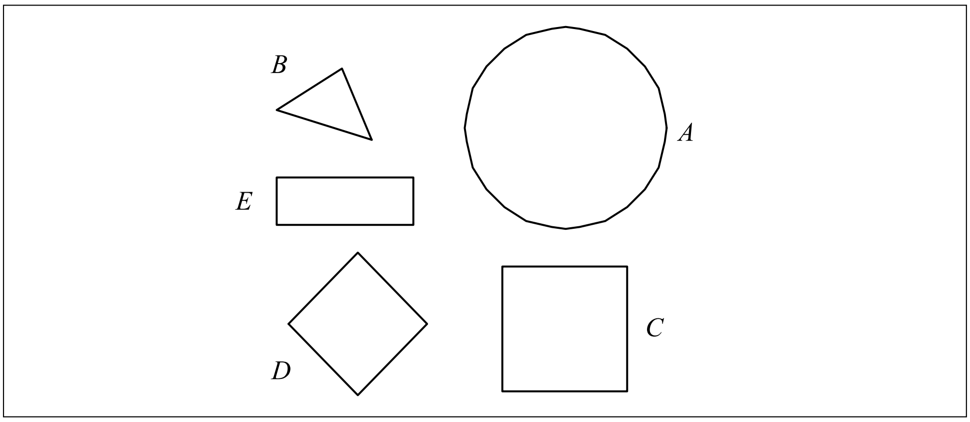

Fig. 1. Estimating the Relative Areas of Five Polygons

to compare them in pairs to reproduce their relative size. First, we arrange the objects we wish to compare among themselves in a square matrix $A = (a_{ij})$ as to their dominance with respect to a common property which is area in this example. The judgments are entered using the fundamental scale derived above to cover a homogeneous range of values from 1–9. An element compared with itself is always assigned the value 1 so the main diagonal entries of the pairwise comparison matrix are all 1. The numbers 3, 5, 7, and 9 correspond to the verbal judgments "moderately more dominant", "strongly more dominant", "very strongly more dominant", and "extremely more dominant" (with 2, 4, 6, and 8 for compromise between these values). Reciprocal values are automatically entered in the transpose position. We are permitted to interpolate values between the integers, if desired.

In the absence of a computer, we can approximate the priorities derived from the matrix by assuming that it is nearly consistent (because in a consistent matrix any column gives the priority vector), normalize each column, and then take the average of the first entries in the five columns, and then the average of the second entries and so on thus obtaining an estimate of the relative areas.

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.

Here is another example (one of many) which shows that the scale works well on homogeneous elements of a real life problem. A matrix of paired comparison judgments is used to estimate relative drink consumption in the United States. To make the comparisons, as shown in table 2 the types of drinks are listed on the left and at the top, and judgment is made as to how strongly the consumption of a drink on the left dominates that of a drink at the top. For example, when coffee on the left is compared with wine at the top, it is thought that it is consumed extremely more and a 9 is entered in the first row and second column position. A 1/9 is automatically entered in the second row and first column position. If the consumption of a drink on the left does not dominate that of a drink at the top, the reciprocal value is entered. For example in comparing coffee and water in the first row and eighth column position, water is consumed more than coffee slightly and a 1/2 is entered. Correspondingly, a value of 2 is entered in the eighth row and first column position. At the bottom of, we see that the derived values and the actual values are close.

T a b l e 2 . Which Drink is Consumed More in the U.S.? An Example of Estimation Using Judgments

Drink Consumption in the U.S.	**Coffee**	**Wine**	**Tea**	**Beer**	**Sodas**	**Milk**	**Water**
Coffee	1	9	5	2	1	1	1/2
Wine	1/9	1	1/3	1/9	1/9	1/9	1/9
Tea	1/5	2	1	1/3	1/4	1/3	1/9
Beer	1/2	9	3	1	1/2	1	1/3
Sodas	1	9	4	2	1	2	1/2
Milk	1	9	3	1	1/2	1	1/3
Water	2	9	9	3	2	3	1

The derived scale based on the judgments in the matrix is:

Coffee	**Wine**	**Tea**	**Beer**	**Sodas**	**Milk**	**Water**
0.177	0.019	0.042	0.116	0.190	0.129	0.327

with a consistency ratio of 0.022.

The actual consumption (from statistical sources) is:

0.180	0.010	0.040	0.120	0.180	0.140	0.330

4. CLUSTERING AND HOMOGENEITY; USING PIVOTS TO EXTEND THE SCALE FROM 1–9 TO 1 – ∞

We note that our ability to make accurate comparisons of widely disparate objects on a common property is limited. We cannot compare with any reliability the very small with the very large. However, we can do it in stages by comparing objects of relatively close magnitudes and gradually increase their sizes until we reach the desired object of large size (see example later). In this process we can think of comparing several close or homogeneous objects for which we obtain a scale of relative values, and then again pairwise compare the next set of larger objects that includes for example the largest object from the previous already compared collection, and then derive a scale for this second set. We then divide all the measurements in the second set by the value of the common object and multiply all the resulting values by the weight of the common element in the first set, thus rendering the two sets to be measurable on the same scale and so on to a third collection of the objects using a common object from the second set.

In Fig. 2 an unripe cherry tomato is eventually and indirectly compared with a large watermelon by first comparing it with a small tomato and a lime, the lime is then used again in a second cluster with a grapefruit and a honey dew where we then divide by the weight of the lime and then multiply by its weight in the first cluster, and then use the honey dew again in a third cluster and so on. In the end we have a comparison of the unripe cherry tomato with the large watermelon and would accordingly extended the scale from 1–9 to 1–721.

Such clustering is essential, and must be done separately for each criterion. We should note that in most decision problems, there may be one or two levels of clusters and conceivably it may go up to three or four adjacent ranges of homogeneous elements (Maslow put them in seven groupings of human needs). Very roughly we have in decreasing order of importance:

1. Survival, health, security, family, friends and basic religious beliefs some people were known to die for;
2. Career, education, productivity and lifestyle;
3. Political and social beliefs and contributions;
4. Beliefs, ideas, and things that are flexible and it does not matter exactly how one advocates or uses them.

Nevertheless one needs them, such as learning to eat with a fork or a chopstick or with the fingers as many people do interchangeably. These categories can be generalized to a group, a corporation, or a government. For very important decisions, two categories may need to be considered. Note that the priorities in two adjacent categories would be sufficiently different, one being an order of magnitude smaller than the other, that in the synthesis, the priorities of the elements in the smaller set have little effect on the decision. How some *undesirable* elements can be compared among themselves and then related to *desirable* ones thus going from negative to positive elements keeping the measurement of the two types positive is done as in the BOCR discussion given later.

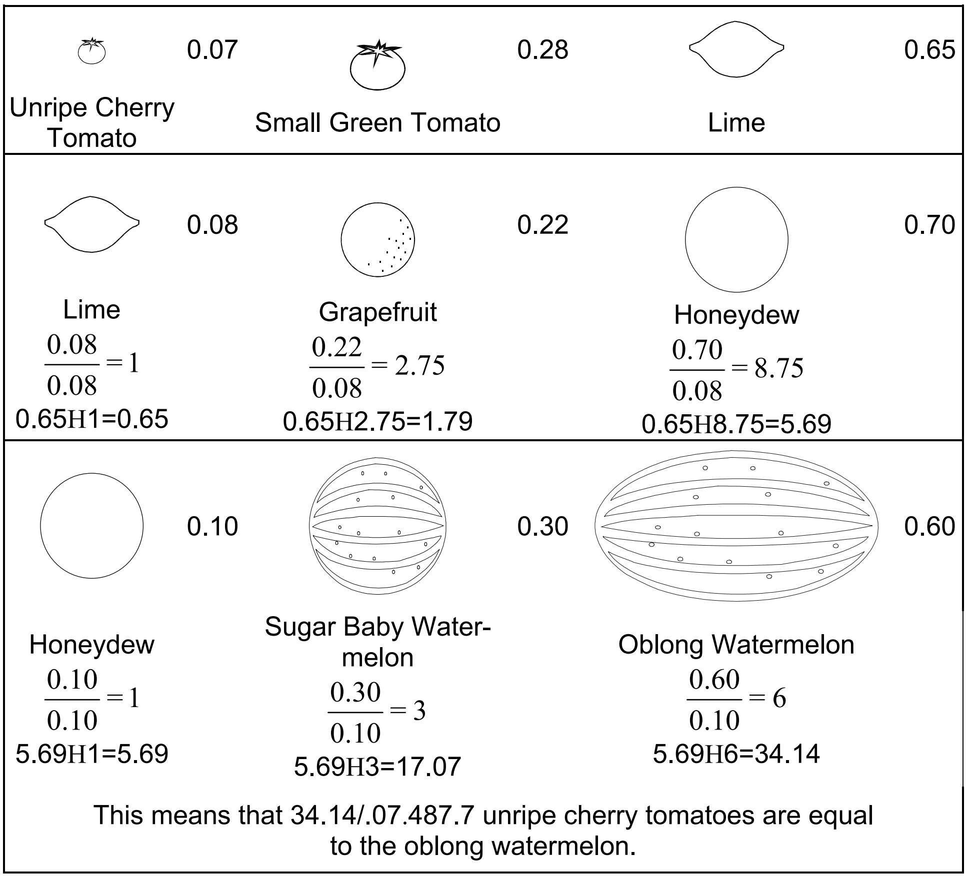

Fig. 2. Comparisons According to Volume

5. WHEN IS A POSITIVE RECIPROCAL MATRIX CONSISTENT?

In light of the foregoing, for the validity of the vector of priorities to describe response, we need greater redundancy and therefore also a large number of comparisons. We now show that for the sake of being close to consistency we need to make a small number, $n(n-1)/2$, of comparisons. An expert may provide n-1 comparisons to fill one row or a spanning tree from which the matrix is consistent and he priorities are easily obtained. So where is the optimum number?

Let us relate the psychological idea of the consistency of judgments and its measurement, to a central concept in matrix theory and also to the size of our channel capacity to process information. It is the principal eigenvalue of a matrix of paired comparisons.

Let $A=[a_{ij}]$ be an *n*-by-*n* positive reciprocal matrix, so all $a_{ii}=1$ and $a_{ij}=1/a_{ji}$ for all $i,j=1,...,n$. Let $w=[w_i]$ be the principal right eigenvector of A, let $D=\text{diag}\,(w_1,...,w_n)$ be the *n*-by-*n* diagonal matrix whose main diagonal entries are the entries of w, and set $E/D^{-1}AD=[a_{ij}\,w_j/w_i]=[\gamma_{ij}]$. Then E is similar to A and is a positive reciprocal matrix since $\gamma_{ji}=a_{ji}\,w_i/w_j=$

$= (a_{ij}\, w_j / w_i)^{-1} = 1/\gamma_{ij}$. Moreover, all the row sums of E are equal to the principal eigenvalue of A:

$$\sum_{j=1}^{n} \varepsilon_{ij} = \sum_j a_{ij} w_j / w_i = [Aw]_i / w_i = \lambda_{\max} w_i / w_i = \lambda_{\max}.$$

The computation

$$n\lambda_{\max} = \sum_{i=1}^{n} (\sum_{j=1}^{n} \varepsilon_{ij}) = \sum_{i=1}^{n} \varepsilon_{ii} + \sum_{\substack{i,j=1 \\ i \neq j}}^{n} (\varepsilon_{ij} + \varepsilon_{ji}) =$$

$$= n + \sum_{\substack{i,j=1 \\ i \neq j}}^{n} (\varepsilon_{ij} + \varepsilon_{ij}^{-1}) \geq n + (n^2 - n)/2 = n^2$$

reveals that $\lambda_{\max} \geq n.$ Moreover, since $x + 1/x \geq 2$ for all $x > 0$, with equality if and only if $x = 1$, we see that $\lambda_{\max} = n$ if and only if all $\gamma_{ij} = 1$, which is equivalent to having all $a_{ij} = w_i / w_j$.

The foregoing arguments show that a positive reciprocal matrix A has $\lambda_{\max} \geq n$, with equality if and only if A is consistent. When A is consistent we have $A^k = n^{k-1} A$ As our measure of deviation of A from consistency, we choose the *consistency index*

$$\mu \equiv \frac{\lambda_{\max} - n}{n - 1}.$$

We have seen that $\mu \geq 0$ and $\mu = 0$ if and only if A is consistent. We can say that as $\mu \to 0,$ $a_{ij} \to w_i / w_j$, or $\varepsilon_{ij} = a_{ij}\, w_j / w_i \to 1$. These two desirable properties explain the term "n" in the numerator of μ; what about the term "$n - 1$" in the denominator? Since trace $(A) = n$ is the sum of all the eigenvalues of A, if we denote the eigenvalues of A that are different from $\lambda_{\max}$ by $\lambda_2, ..., \lambda_{n-1}$, we see that $n = \lambda_{\max} + \sum_{i=2}^{n} \lambda_i$, so $n - \lambda_{\max} = \sum_{i=2}^{n} \lambda_i$ and $\mu = $ $= \frac{1}{n-1} \sum_{i=1}^{n} \lambda_i$ is the average of the non-principal eigenvalues of A.

It is an easy, but instructive, computation to show that $\lambda_{\max} = 2$ for every 2-by-2 positive reciprocal matrix:

$$\begin{bmatrix} 1 & \alpha \\ \alpha^{-1} & 1 \end{bmatrix} \begin{bmatrix} 1 + \alpha \\ (1 + \alpha)\alpha^{-1} \end{bmatrix} = 2 \begin{bmatrix} 1 + \alpha \\ (1 + \alpha)\alpha^{-1} \end{bmatrix}.$$

Thus, every 2-by-2 positive reciprocal matrix is consistent.

Not every 3-by-3 positive reciprocal matrix is consistent, but in this case we are fortunate to have again explicit formulas for the principal eigenvalue and eigenvector. For

$$A = \begin{bmatrix} 1 & a & b \\ 1/a & 1 & c \\ 1/b & 1/c & 1 \end{bmatrix}$$

we have $\lambda_{\max} = 1 + d + d^{-1}$, $d = (ac/b)^{1/3}$ and

$$w_1 = bd \Big/ \left(1 + bd + \frac{c}{d}\right),\ w_2 = c \Big/ d\left(1 + bd + \frac{c}{d}\right),\ w_3 = 1 \Big/ \left(1 + bd + \frac{c}{d}\right).$$

Note that $\lambda_{\max} = 3$ when $d = 1$ or $c = b/a$, which is true if and only if A is consistent.

In order to get some feel for what the consistency index might be telling us about a positive n-by-n reciprocal matrix A, consider the following simulation: choose the entries of A above the main diagonal at random from the 17 values $\{1/9, 1/8, \ldots, 1, 2, \ldots, 8, 9\}$. Then fill in the entries of A below the diagonal by taking reciprocals. Put ones down the main diagonal and compute the consistency index. Do this 50000 times and take the average, which we call the *random index*. Table 3 shows the values obtained from one set of such simulations and also their first order differences, for matrices of size 1, 2,…,15.

Table 3. Random Index

Order	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
R.I.	0	0	0.52	0.89	1.11	1.25	1.35	1.40	1.45	1.49	1.52	1.54	1.56	1.58	1.59
First Order Differences		0	0.52	0.37	0.22	0.14	0.10	0.05	0.05	0.04	0.03	0.02	0.02	0.02	0.01

Fig. 3 below is a plot of the first two rows of Table 3. It shows the asymptotic nature of random inconsistency. We also learn that one should not compare more than about seven elements because the inconsistency increases due to a correct mental limitation, I believe is a result of evolution, to deal with a few things at a time [4].

Since it would be pointless to try to discern any priority ranking from a set of random comparison judgments, we should probably be uncomfortable about proceeding unless the consistency index of a pairwise comparison matrix is very much smaller than the corresponding random index value in Table 3. The *consistency ratio* (C.R.) of a pairwise comparison matrix is the ratio of its consistency index :, defined above, to the corresponding random index value in Table 3. As a rule of thumb, we do not recommend proceeding if the consistency ratio is more than about .10 for $n \geq 4$. For $n = 3$, we recommend that the $C.R.$ be less than 0.05. By way of further elaboration, inconsistency may be thought of as an adjustment needed to improve the consistency of the comparisons. But the adjustment should not be as large as the judgment itself, nor so small that using it is of no consequence. Thus inconsistency should be just one order of magnitude smaller. On a scale from zero to one, the overall inconsistency should be around 10%. The requirement of 10% cannot be made smaller such as 1% or 0.1% without trivializing the impact of inconsistency. But inconsistency itself is important because without it, new knowledge that changes preference cannot be admitted. Assuming that all knowledge should be consistent contradicts experience that requires continued revision of understanding.

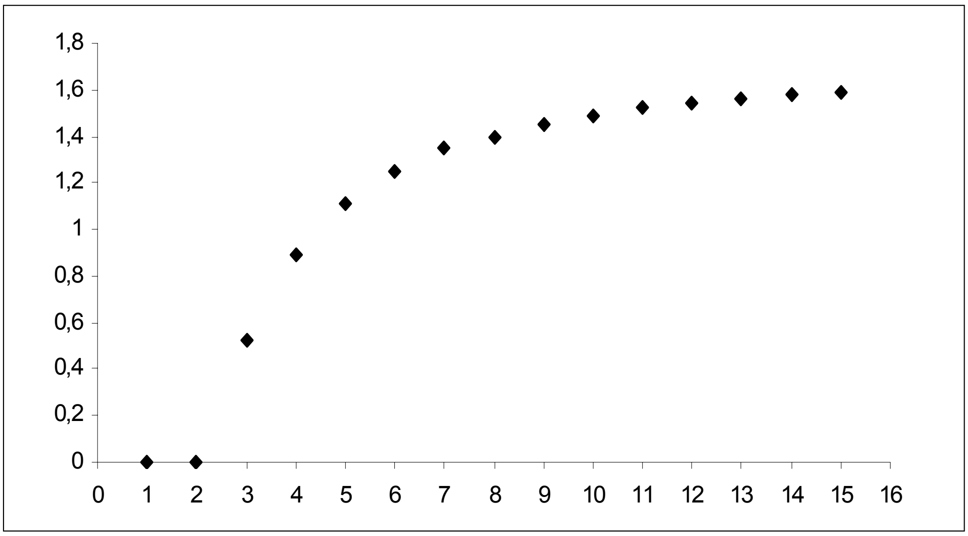

Fig. 3. Plot of Random Inconsistency

If the $C.R.$ is larger than desired, we do three things:

1) find the most inconsistent judgment in the matrix (for example, that judgment for which $\varepsilon_{ij} = a_{ij} w_j / w_i$ is largest);

2) determine the range of values to which that judgment can be changed corresponding to which the inconsistency would be improved;

3) ask the judge to consider, if he can, change his judgment to a plausible value in that range. If he is unwilling, we try with the second most inconsistent judgment and so on. If no judgment is changed the decision is postponed until better understanding of the stimuli is obtained.

Before proceeding further, the following observations may be useful for a better understanding of the importance of the concept of a limit on our ability to process information and also change in information. The quality of response to stimuli is determined by three factors. Accuracy or validity, consistency, and efficiency or amount of information generated. Our judgment is much more sensitive and responsive to large perturbations. When we speak of perturbation, we have in mind numerical change from consistent ratios obtained from priorities. The larger the inconsistency and hence also the larger the perturbations in priorities, the greater is our sensitivity to make changes in the numerical values assigned. Conversely, the smaller the inconsistency, the more difficult it is for us to know where the best changes should be made to produce not only better consistency but also better validity of the outcome. Once near consistency is attained, it becomes uncertain which coefficients should be perturbed by small amounts to transform a near consistent matrix to a consistent one. If such perturbations were forced, they could be arbitrary and thus distort the validity of the derived priority vector in representing the underlying decision.

The third row of Table 3 gives the differences between successive numbers in the second row. Fig. 4 is a plot of these differences and shows the importance of the number **seven** as a cutoff point beyond which the differences are less than 0.10 where we are not sufficiently sensitive to make accurate changes in judgment on several elements simultaneously.

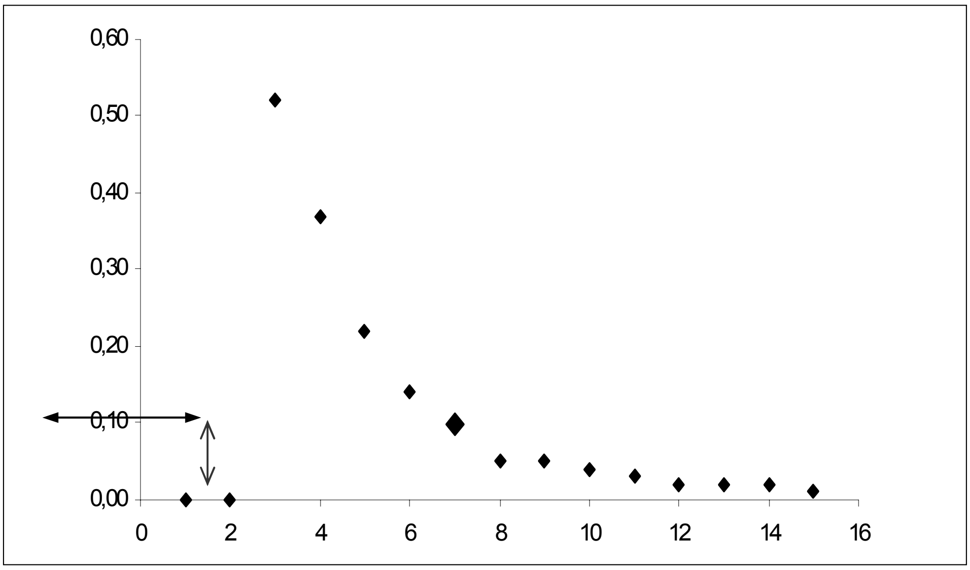

Fig. 4. Plot of First Differences in Random Inconsistency

6. IN THE ANALYTIC HIERARCHY PROCESS ADDITIVE COMPOSITION IS NECESSARY

Synthesis in the AHP involves weighting the priorities of elements compared with respect to an element in the next higher level, called a parent element, by the priority of that element and adding over all such parents for each element in the lower level. In the AHP we assume that the criteria are preferentially independent (inner independent among themselves) and therefore use additive synthesis. This process of weighting and adding is carried out from the top to the bottom of a hierarchy. The result is a multilinear form involving sums of products of all the priorities from the top to the bottom. Multilinear forms are important for representing the weights of all the factors involved in the prioritization process to capture nonlinear effects as accurately as desired by expanding the hierarchy both laterally and in depth to include all the important factors considered to have an influence on the outcome. When the criteria are not independent, their weights are derived as a function of the weights of the other criteria. The resulting weights are used as if they are preferentially independent. Additive composition is needed to make possible tradeoff among them. The non-additive component is already subsumed in the way we obtain the weights of these criteria.

Let us consider the example of two criteria and three alternatives measured in the same scale such as dollars. If the criteria are each assigned the value 1, then weighting and adding produces the correct dollar value. However, it does not if the weights of the criteria are normalized, with each criterion having a weight of 0.5. Once the criteria are given in relative terms, so must the alternatives be and then the criteria must be related according to importance. A criterion that measures values in pennies cannot be as important as another measured in thousands of dollars. In this case, the only meaningful importance of a criterion is the ratio of the total money for the alternatives under it to the total money for the alternatives under both criteria. By using these weights for the criteria, rather than 0.5 and 0.5, one obtains the correct final relative values for the alternatives. The AHP is a special case of the Analytic Network Process. As we see shall see later, the ANP automatically assigns the criteria the correct weights, if one only uses the normalized values of the alternatives under each criterion and also the normalized values for each alternative under all the criteria without any special attention to weighting the criteria. Its operation of raising the supermatrix to the limit powers does that. The ANP is based on additive synthesis. This is an important example that shows that additive synthesis in both the AHP and ANP is necessary to obtain the correct final result in relative terms when the weights of the criteria depend on the weights of the alternatives.

An Investment Example

An individual has three alternate ways, A_1, A_2 and A_3, of investing a sum of money for the same period of time. There are two types of returns, C_1 and C_2 (for example, capital appreciation and interest), as shown in Table 4. The question is, which is the best investment to make in terms of actual dollars earned?

The returns for each investment are shown below. It is easy to calculate the actual total cost for each house by simply adding the two numbers; the relative cost is then obtained by normalizing as shown in Table 4 below. We will then show that the ANP process of prioritizing the houses in terms of the criteria and then the criteria in terms of the houses will give the same relative costs as the arithmetic computation. This validates the ANP as an additive synthesis process with a real example where measureable criteria are involved.

Table 4. Calculating Returns Arithmetically

Alternatives	**Criterion C_1** Unnormalized weight = 1.0	**Criterion C_2** Unnormalized weight = 1.0	Weighted Sum Unnormalized	Normalized or Relative values
A_1	200	150	350	350/1300=0.269
A_2	300	50	350	350/1300=0.269
A_3	500	100	600	600/1300=0.462
Column totals	1000	300	1300	1

What is the relative importance of each criterion? We cannot simply assign each criterion the value one because the values assigned to the criteria need to be normalized yielding 0.5, 0.5. But normalization indicates relative importance. Relative values require that criteria be examined as to their relative importance with respect to each other. What is the relative importance of a criterion, or what numbers should the criteria be assigned that reflect their relative importance? Weighting each criterion by the proportion of the resource under it, as shown in Table 5, and multiplying and adding as in the additive synthesis of the AHP, we get the same correct answer obtained in Table 5 using arithmetic. For criterion C_1 we have $(200+300+500)/[(200+300+500)+(150+50+100)]=1000/1300$ and for criterion C_2 we have $(150+50+100)/[(200+300+500)+(150+50+100)]=300/1300$. Here the criteria are normalized accordingly, and their weights sum to one. We see that when the criteria are normalized, the alternatives must also be normalized to get the right answer. For example if we look in Table 4 we have 350/1300 for the priority of alternative A_1. Now if we simply weight and add the values for alternative A_1 in Table 4 we get for its final relative value $200(1000/1300)+150(300/1300)=350/1300$ as it should be from Table 4. However if we normalize as in Table 5 below we get $(200/1000)(1000/1300)+(150/300)(300/1300)=350/1300$. We see that if the priorities of the alternatives are not normalized one does not get meaningful answers. Thus at least in this case, normalization of the priorities of the alternatives is necessary when the priorities of the criteria depend on the priorities of the alternatives.

Table 5. Normalized Criteria Weights and Normalized Alternative Weights From Measurements in Same Scale (Additive Synthesis)

Alternatives	**Criterion C_1** Normalized weight = 1000/1300=0.7692	**Criterion C_2** Normalized weight = 300/1300=0.2308	Weighted Sum
A_1	200/1000	150/300	350/1300 = 0.2692
A_2	300/1000	50/300	350/1300 = 0.2692
A_3	500/1000	100/300	600/1300 =0.4615

7. THE DISTRIBUTIVE AND IDEAL MODES

We have seen in the example of the last section that in order to obtain the correct final relative values for the alternatives when measurements on a measurement scale are given, it is essential that the priorities of the criteria be derived from the priorities of the alternatives. This is followed by using the normalized weights of the alternatives multiplied by the priorities of the criteria and summed over the criteria to obtain the relative value for each alternative. Thus when the criteria depend on the alternatives we need to normalize the values of the alternatives to obtain the final result. This procedure is known as the distributive mode of the AHP. It is also used in case of functional dependence of the criteria on the alternatives for example. Here the criteria are compared with respect to each alternative separately as in the ANP. For example one asks about a given individual: is this individual better as a musician or as a teacher and how much better? The dominant mode of synthesis in the ANP is the distributive mode.

When the criteria do not depend on the values of the alternatives we need to derive their priorities by comparing them pairwise with each other with respect to higher level criteria or the goal. It is a process of trading off one unit of one criterion against a unit of another, an ideal alternative from one against an ideal alternative from another. To determine the ideal the alternatives are divided by the largest value among them for each criterion. In that case the process of weighting and adding assigns each of the remaining alternatives a value that is proportionate to the value 1 given to the highest rated alternative. In this way the alternatives are weighted by the priorities of the criteria and summed to obtain the weights of the alternatives. This is the ideal mode of the AHP. Thus the distributive mode is essential for synthesizing the weights of alternatives with respect to tangible criteria with the same scale of measurement into a single criterion for that scale and then they are treated as intangibles and compared pairwise and combined with other intangibles with the ideal mode. The dominant mode of synthesis in the AHP where the criteria are independent from the alternatives is the ideal mode. The standard mode for synthesizing in the ANP where criteria depend on alternatives and also alternatives may depend on other alternatives is the distributive mode.

8. RATING ALTERNATIVES ONE AT A TIME IN THE AHP

The AHP has a second way to derive priorities known as absolute measurement. It involves making paired comparisons but the criteria just above the alternatives, known as the covering criteria, are assigned intensities that vary in number and type. For example they can simply be: high, medium and low; or they can be: excellent, very good, good, average, poor and very poor; or for experience: more than 15 years, between 10 and 15, between 5 and 10 and less than 5 and so on. These intensities are compared pairwise among themselves to obtain their priorities as to importance, and they are then put in ideal form by dividing by the largest value. Finally each alternative is assigned an intensity with its accompanying priority (called rating) for each criterion. The priority of each intensity is weighted by the priority of its criterion and summed over the weighted intensities for each alternative to obtain that alternative's final rating that also belong to a ratio scale. It is often necessary to have categories of ratings for alternatives that are widely disparate so that one can rate the alternatives correctly. Ratings are useful when standards are established with which the alternatives must comply. They are also useful when the number of alternatives n is very large to perform pairwise comparisons on them for each criterion. In this case if the number of criteria is c, the number of rating operations in rating the alternatives is cn, whereas doing all the pairwise judgments involves $cn(n-1)/2$ comparisons.

9. NEGATIVE PRIORITIES IN THE AHP

Were one to subtract the values of the alternatives under a criterion, taken as costs, from those under another, taken as gains, both measured in the same unit, and then normalize the results, the outcome can again be duplicated in relative terms. We total each of the two columns for the alternatives under the two criteria, subtract the second total from the first. We then assign each criterion the sum of the values of the alternatives under it divided by the absolute value of the foregoing difference (treating a difference of zero as a special case in which benefits and costs have equal weights). Note that the criteria are normalized with respect to differences rather than sums. We then weight by the priorities of the criteria and add. The outcome, as in the example in Table 5, is identical to subtracting the second value for each alternative from the first summing and then dividing by the absolute value of the sum of these differences. Clearly the process requires that we use the distributive and not the ideal mode. The following example speaks for itself. The original values were the numerators of the two middle columns of Table 6.

In the absence of a unit, at first it is not obvious as how to combine positive and negative numbers as priorities. Thus it is not possible for example to compare the relative importance of benefits with the relative importance of costs. Pleasure is not a higher form of pain nor is good a higher form of evil. The good and the bad are different but they are opposites of the same type or dimension of measurement. One way that has been justifiably used in the AHP instead of negative priorities is to take reciprocals, weight and then add them to other positive priorities.

Table 6. Subtracting Priorities From Same Scale Measurements

Alternatives	Criterion C_1 Normalized weight = 10/4	Criterion C_2 Normalized weight = 14/4	Weighted difference
A_1	3/10	8/14	–5/4
A_2	2/10	1/14	1/4
A_3	5/10	5/14	0

From the requirement of dominance by using a unit in paired comparisons, we know that we can only ask how much an element dominates another element and not how much an element is dominated by another element. It is not meaningful to do it the opposite way without first using the smaller element as the unit to determine how many times larger is the more dominant element and then estimating the smaller one as a fraction of it. Thus not only does one ask how much more important one element is than another according to benefits and opportunities, but also how much more costly or risky one element is than another with respect to a certain criterion.

We know that people can trade off the benefits of an alternative against its costs in making a decision but they do not do it by a process of wholesale comparison. They do it with respect to their own satisfaction (strategic) criteria and by rating the contribution of benefits and the costs of that alternative separately to the fulfillment of those criteria. When there are several alternatives, one uses for each of the benefits and costs that alternative with the largest composite ideal priority. It may be that the best-ranked alternative under benefits differs from that under costs, but in any case one uses the ideal alternative in developing the ratings for the benefits. The same approach applies to the costs. The results obtained from the ratings take the form of non-normalized priorities for the benefits and the costs. Normalizing these rating values yields the desired priorities that enable us to tradeoff the benefits and costs of all the alternatives. The example in the next section will help clarify these ideas.

More generally, in many decision problems four kinds of concerns are considered: benefits, opportunities, costs and risks; which we abbreviate as **BOCR**. The first two are advantageous and hence are positive and the second two are disadvantageous and are therefore negative [5]. We have sometimes justifiably kept the last two positive in a situation where the decision was already made for example to buy a car, and low cost as determined by the normalized reciprocal costs of the alternatives was seen as a benefit that is then weighted and added to the benefits.

Another and a more accurate way to deal with **BOCR** is to realize that through normalization of the principal eigenvector one obtains a dimensionless set of numbers that belong to an absolute scale. It is known that absolute numbers can be both positive and negative and hence it is not necessary to confine the **BOCR** to being positive.

There are at least four ways to combine **BOCR** priorities with corresponding normalized weights b, o, c, r obtained by rating **B** and then **C** and then **O** and finally **R** separately. The first way is the traditional one in which weighting amounts to multiplying by the same constant. They are:

BO/CR,

$b\mathbf{B} + o\mathbf{O} + c(1/\mathbf{C}) + r(1/\mathbf{R})$,

$b\mathbf{B} + o\mathbf{O} + c(1-\mathbf{C}) + r(1-\mathbf{R})$,

$b\mathbf{B} + o\mathbf{O} - c\,\mathbf{C} - r\mathbf{R}$.

The four methods do not always yield the same best answer. One counter example demonstrates the truth of this observation. The question now is how to interpret these priorities and use them appropriately in different situations. The first is a tradeoff between a unit of **BO** against a unit of **CR**, a unit of the desirable against a unit of the undesirable. The second is a sum of the advantages obtained when committed to action with low values of the disadvantages (the lesser of the evils) considered as good or positive. The third is more optimistic and considers the residual or complementary value, the fact that «not all is bad» as a positive measure. The fourth and last simply subtracts the sum of the weighted «bads» from the sum of the weighted «goods» and can give rise to negative priorities. Below we give an example to illustrate these four ways of aggregating **BOCR**. Which to use depends on the circumstances one faces. The first is used when resources are limited.

This paper will be continued in Part 2.2.

REFERENCES

1. *Saaty, Thomas L.* The Analytic Network Process, RWS Publications, 4922 Ellsworth Avenue, Pittsburgh, Pa. 15213, 2001.
2. *Saaty, Rozann W.* Decision Making in Complex Environments: The Analytic Network Process (ANP) for Dependence and Feedback; a manual for the ANP Software; Creative Decisions Foundation, 4922 Ellsworth Avenue, Pittsburgh, Pa. 15213, 2002.
3. *Saaty, Thomas L.* Fundamentals of the Analytic Hierarchy Process, RWS Publications, 4922 Ellsworth Avenue, Pittsburgh, Pa. 15413, 2000.
4. *Saaty Thomas L.* and Mujgan S. Ozdemir. Why the Magic Number Seven Plus or Minus Two, (Submitted for publication.)

Received 25.01.2003