JOURNAL OF MATHEMATICAL PSYCHOLOGY 15, 234–281 (1977)

A Scaling Method for Priorities in Hierarchical Structures

THOMAS L. SAATY

University of Pennsylvania, Wharton School, Philadelphia, Pennsylvania 19174

The purpose of this paper is to investigate a method of scaling ratios using the principal eigenvector of a positive pairwise comparison matrix. Consistency of the matrix data is defined and measured by an expression involving the average of the nonprincipal eigenvalues. We show that $\lambda_{\max} = n$ is a necessary and sufficient condition for consistency. We also show that twice this measure is the variance in judgmental errors. A scale of numbers from 1 to 9 is introduced together with a discussion of how it compares with other scales. To illustrate the theory, it is then applied to some examples for which the answer is known, offering the opportunity for validating the approach. The discussion is then extended to multiple criterion decision making by formally introducing the notion of a hierarchy, investigating some properties of hierarchies, and applying the eigenvalue approach to scaling complex problems structured hierarchically to obtain a unidimensional composite vector for scaling the elements falling in any single level of the hierarchy. A brief discussion is also included regarding how the hierarchy serves as a useful tool for decomposing a large-scale problem, in order to make measurement possible despite the now-classical observation that the mind is limited to 7 ± 2 factors for simultaneous comparison.

1. INTRODUCTION

A fundamental problem of decision theory is how to derive weights for a set of activities according to importance. Importance is usually judged according to several criteria. Each criterion may be shared by some or by all the activities. The criteria may, for example, be objectives which the activities have been devised to fulfill. This is a process of multiple criterion decision making which we study here through a theory of measurement in a hierarchical structure.

The object is to use the weights which we call priorities, for example, to allocate a resource among the activities or simply implement the most important activities by rank if precise weights cannot be obtained. The problem then is to find the relative strength or priorities of each activity with respect to each objective and then compose the result obtained for each objective to obtain a single overall priority for all the activities. Frequently the objectives themselves must be prioritized or ranked in terms of yet another set of (higher-level) objectives. The priorities thus obtained are then used as weighting factors for the priorities just derived for the activities. In many applications we have noted that the process has to be continued by comparing the higher-level objectives in terms of still higher ones and so on up to a single overall objective. (The top level need not have a single element in which case one would have to assume rather than derive weights for the elements in that level.) The arrangement of the activities;

ISSN 0022-2496

first set of objectives, second set, and so on to the single element objective defines a hierarchical structure.

The paper is concerned with developing a method for scaling the weights of the elements in each level of the hierarchy with respect to an element (e.g. criterion or objective) of the next higher level. We construct a matrix of pairwise comparisons of the activities whose entries indicate the strength with which one element dominates another as far as the criterion with respect to which they are compared is concerned.

If, for example, the weights are w_i, $i = 1,..., n$, where n is the number of activities, then an entry a_{ij} is an estimate of w_i/w_j. This scaling formulation is translated into a largest eigenvalue problem. The Perron–Frobenius theory (Gantmacher, 1960) ensures the existence of a largest real positive eigenvalue for matrices with positive entries whose associated eigenvector is the vector of weights. This vector is normalized by having its entries sum to unity. It is unique.

Thus the activities in the lowest level have a vector of weights with respect to each criterion in the next level derived from a matrix of pairwise comparisons with respect to that criterion.

The weight vectors at any one level are combined as the columns of a matrix for that level. The weight matrix of a level is multiplied on the right by the weight matrix (or vector) of the next higher level. If the highest level of the hierarchy consists of a single objective, then these multiplications will result in a single vector of weights which will indicate the relative priority of the entities of the lowest level for accomplishing the highest objective of the hierarchy. If one decision is required, the option with the highest weight is selected; otherwise, the resources are distributed to the options in proportion to their weights in the final vector. Other optimization problems with constraints have been considered elsewhere.

Special emphasis is placed in this work on the integration of human judgments into decisions and on the measurement of the consistency of judgments. From a theoretical standpoint consistency is a necessary condition for representing a real-life problem with a scale; however, it is not sufficient. The actual validation of a derived scale in practice rests with statistical measures, with intuition, and with pragmatic justification of the results.

2. Ratio Scales from Reciprocal Pairwise Comparison Matrices

Suppose we wish to compare a set of n objects in pairs according to their relative weights (assumed to belong to a ratio scale). Denote the objects by $A_1,..., A_n$ and their weights by $w_1,..., w_n$. The pairwise comparisons may be represented by a matrix as follows:

$$A = \begin{array}{c|cccc} & A_1 & A_2 & \cdots & A_n \\ \hline A_1 & w_1/w_1 & w_1/w_2 & \cdots & w_1/w_n \\ A_2 & w_2/w_1 & w_2/w_2 & \cdots & w_2/w_n \\ \vdots & \vdots & \vdots & & \vdots \\ A_n & w_n/w_1 & w_n/w_2 & & w_n/w_n \end{array}$$

This matrix has positive entries everywhere and satisfies the reciprocal property $a_{ji} = 1/a_{ij}$. It is called a reciprocal matrix. We note that if we multiply this matrix by the transpose of the vector $w^T \equiv (w_1, \ldots, w_n)$ we obtain the vector nw.

Our problem takes the form

$$Aw = nw.$$

We started out with the assumption that w was given. But if we only had A and wanted to recover w we would have to solve the system $(A - nI)w = 0$ in the unknown w. This has a nonzero solution if and only if n is an eigenvalue of A, i.e., it is a root of the characteristic equation of A. But A has unit rank since every row is a constant multiple of the first row. Thus all the eigenvalues λ_i, $i = 1, \ldots, n$, of A are zero except one. Also, it is known that

$$\sum_{i=1}^{n} \lambda_i = \operatorname{tr}(A) \equiv \text{sum of the diagonal elements} = n.$$

Therefore only one of the λ_i, which we call $\lambda_{\max}$, equals n; and

$$\lambda_i = 0, \qquad \lambda_i \neq \lambda_{\max}.$$

The solution w of this problem is any column of A. These solutions differ by a multiplicative constant. However, it is desirable to have this solution normalized so that its components sum to unity. The result is a unique solution no matter which column is used. We have recovered the scale from the matrix of ratios.

The matrix A satisfies the "cardinal" consistency property $a_{ij}a_{jk} = a_{ik}$ and is called consistent. For example if we are given any row of A, we can determine the rest of the entries from this relation. This also holds for any set of n entries whose graph is a spanning cycle of the graph of the matrix.

Now suppose that we are dealing with a situation in which the scale is not known but we have estimates of the ratios in the matrix. In this case the cardinal consistency relation (elementwise dominance) above need not hold, nor need an ordinal relation of the form $A_i > A_j$, $A_j > A_k$ imply $A_i > A_k$ hold (where the A_i are rows of A).

As a realistic representation of the situation in preference comparisons, we wish to account for inconsistency in judgments because, despite their best efforts, people's feelings and preferences remain inconsistent and intransitive.

We know that in any matrix, small perturbations in the coefficients imply small perturbations in the eigenvalues. Thus the problem $Aw = nw$ becomes $A'w' = \lambda_{\max}w'$. We also know from the theorem of Perron–Frobenius that a matrix of positive entries has a real positive eigenvalue (of multiplicity 1) whose modulus exceeds those of all other eigenvalues. The corresponding eigenvector solution has nonnegative entries and when normalized it is unique. Some of the remaining eigenvalues may be complex.

Suppose then that we have a reciprocal matrix. What can we say about an overall estimate of inconsistency for both small and large perturbations of its entries? In other words how close is $\lambda_{\max}$ to n and w' to w? If they are not close, we may either revise the estimates in the matrix or take several matrices from which the solution vector w'

may be improved. Note that improving consistency does not mean getting an answer closer to the "real" life solution. It only means that the ratio estimates in the matrix, as a sample collection, are closer to being logically related than to being randomly chosen.

From here on we use $A = (a_{ij})$ for the estimated matrix and w for the eigenvector. There should be no confusion in dropping the primes.

It turns out that a reciprocal matrix A with positive entries is consistent if and only if $\lambda_{\max} = n$ (Theorem 1 below). With inconsistency $\lambda_{\max} > n$ always. One can also show that ordinal consistency is preserved, i.e., if $A_i \geqslant A_j$ (or $a_{ik} \geqslant a_{jk}$, $k = 1,..., n$) then $w_i \geqslant w_j$ (Theorem 2 below). We now establish $(\lambda_{\max} - n)/(n - 1)$ as a measure of the consistency or reliability of information by an individual to be of the form w_i/w_j. We assume that because of possible error the estimate has the form $w_i/w_j\, \epsilon_{ij}$ where $\epsilon_{ij} > 0$.

First we note that to study the sensitivity of the eigenvector to perturbations in a_{ij} we cannot make a precise statement about a perturbation $dw = (dw_1 ,..., dw_n)$ in the vector $w = (w_1 ,..., w_n)$ because everywhere we deal with w, it appears in the form of ratios w_i/w_j or with perturbations (mostly multiplicative) of this ratio. Thus, we cannot hope to obtain a simple measure of the absolute error in w.

From general considerations one can show that the larger the order of the matrix the less significant are small perturbations or a few large perturbations on the eigenvector. If the order of the matrix is small, the effect of a large array perturbation on the eigenvector can be relatively large. We may assume that when the consistency index shows that perturbations from consistency are large and hence the result is unreliable, the information available cannot be used to derive a reliable answer. If it is possible to improve the consistency to a point where its reliability indicated by the index is acceptable, i.e., the value of the index is small (as compared with its value from a randomly generated reciprocal matrix of the same order), we can carry out the following type of perturbation analysis.

The choice of perturbation most appropriate for describing the effect of inconsistency on the eigenvector depends on what is thought to be the psychological process which goes on in the individual. Mathematically, general perturbations in the ratios may be reduced to the multiplicative form mentioned above. Other perturbations of interest can be reduced to the general form $a_{ij} = (w_i/w_j)\, \epsilon_{ij}$. For example,

$$(w_i/w_j) + \alpha_{ij} = (w_i/w_j)(1 + (w_j/w_i)\, \alpha_{ij}).$$

Starting with the relation

$$\lambda_{\max} = \sum_{j=1}^{n} a_{ij}(w_j/w_i),$$

from the ith component of $Aw = \lambda_{\max} w$, we consider the two real-valued parameters $\lambda_{\max}$ and μ, the average of λ_i, $i \geqslant 2$ (even though they can occur as complex conjugate numbers),

$$\mu = -(1/(n-1)) \sum_{i=2}^{n} \lambda_i = (\lambda_{\max} - n)/(n - 1) \geqslant 0, \qquad \lambda_{\max} \equiv \lambda_1$$

It is desired to have μ near zero, thus also to have $\lambda_{\max}$, which is always $\geqslant n$, near its lower bound n, and thereby obtain consistency. Now we show that $(\lambda_{\max} - n)/(n-1)$ is related to the statistical root mean square error. To see this, we have from

$$\lambda_{\max} - 1 = \sum_{j \neq i} a_{ij} \frac{w_j}{w_i},$$

that

$$n\lambda_{\max} - n = \sum_{1 \leqslant i < j \leqslant n} a_{ij} \frac{w_j}{w_i} + a_{ji} \frac{w_i}{w_j},$$

and therefore

$$\mu = \frac{\lambda_{\max} - n}{n-1} = \frac{1}{n-1} - \frac{n}{n-1} + \frac{1}{n(n-1)} \sum_{1 \leqslant i < j \leqslant n} a_{ij} \frac{w_j}{w_i} + a_{ji} \frac{w_i}{w_j}.$$

Let $a_{ij} = (w_i/w_j)\,\epsilon_{ij}$, $\epsilon_{ij} > 0$. Clearly, we have consistency at $\epsilon_{ij} = 1$. Now by imposing the reciprocal relation $a_{ji} = 1/a_{ij}$, we have:

$$\mu = -1 + \frac{1}{n(n-1)} \sum_{1 \leqslant i < j \leqslant n} \left(\epsilon_{ij} + \frac{1}{\epsilon_{ij}}\right),$$

which $\to 0$ as $\epsilon_{ij} \to 1$. Also, μ is convex in the ϵ_{ij} since $\epsilon_{ij} + (1/\epsilon_{ij})$ is convex (and has its minimum at $\epsilon_{ij} = 1$), and the sum of convex functions is convex. Thus, μ is small or large depending on ϵ_{ij} being near or far from unity, respectively; i.e., near or far from consistency.

If we write $\epsilon_{ij} = 1 + \delta_{ij}$, we have

$$\mu = (1/n(n-1)) \sum_{1 \leqslant i < j \leqslant n} \delta_{ij}^2 - (\delta_{ij}^3/1 + \delta_{ij}).$$

Let us assume that $|\,\delta_{ij}\,| < 1$ (and hence that $\delta_{ij}^3/(1+\delta_{ij})$ is small compared with δ_{ij}^2). This is a reasonable assumption for an unbiased judge who is limited by the "natural" greatest lower bound -1 on δ_{ij} (since a_{ij} must be greater than zero) and would tend to estimate symmetrically about zero in the interval $(-1, 1)$. Now, $\mu \to 0$ as $\delta_{ij} \to 0$. Multiplication by 2 gives the variance of the δ_{ij}. Thus, 2μ is this variance.

Suppose now we wish to develop a test of a hypothesis of consistency. Perfect consistency is stated in the null hypothesis as

$$H_0 : \mu = 0.$$

We test it versus its logical one-sided alternative

$$H_1 : \mu > 0.$$

The appropriate test statistic is

$$m = (\tilde{\lambda}_{\max} - n)/(n-1),$$

where $\tilde{\lambda}_{\max}$ is the maximum observed eigenvalue of the matrix whose elements, a_{ij}, contain random error. Developing a statistical measure for consistency requires finding the distribution of the statistic, m. While its specific form is beyond the scope of this paper, we observe that m follows a nonnegative probability distribution whose variance is twice its mean $\bar{x}$ and appears to be quite similar to the χ^2 distribution if we assume that all δ_{ij} are $N(0, \sigma^2)$ on $(-1, 1)$. Analytically one may have to experiment with other distributions such as the β distribution.

For our purposes, without knowing the distribution, we use the conventional ratio $(\bar{x} - \mu_0)/(2\bar{x})^{1/2}$ with $\mu_0 = 0$, i.e., we use $(\bar{x}/2)^{1/2}$ in a qualitative test to confirm the null hypothesis when the test statistic is, say, $\leqslant 1$. Thus when $\bar{x} > 2$ it is possible that inconsistency is indicated.

There are several advantages of the eigenvalue method in developing a ratio scale as compared with direct estimates of the scale or with least-square methods. For example, when compared with the former, it captures more information through redundancy of information obtained from pairwise comparisons and the use of reciprocals. When compared with either method, it addresses the question of the consistency by a single numerical index and points to the reliability of the data and to revisions in the matrix.

There is no easy way to study the sensitivity of the eigenvector w to errors in A. Apart from experiments and the many illustrations, particularly when the order of the matrix is large, one may use the following formula, complicated because of the many calculations it entails (Wilkinson, 1965):

$$\Delta w_1 = \sum_{j=2}^{n} (v_j^T(\Delta A)\, w_1/(\lambda_1 - \lambda_j)\, v_j^T w_j)\, w_j\,, \qquad w_1 \text{ corresponds to } \lambda_{\max}\,.$$

Note that this equation requires the computation of the eigenvalues λ_i, $i = 1, 2,..., n$, with $\lambda_1 = \lambda_{\max}$, the right and left eigenvectors of A, w_i, and v_i, $i = 1, 2,..., n$. We have shown that w_i is generally insensitive to small perturbations in A for our approach, since near consistency λ_1 is well separated from λ_j and $v_j^T w_j$ is never arbitrarily small.

As already mentioned, it is easy to prove that the solution of the problem $Aw = nw$ when A is consistent is given by the normalized row sums or any normalized column of A. In addition, the solution to $Aw = \lambda_{\max} w$ when $\lambda_{\max}$ is close to n may be approximated by normalizing each column of A and taking the average over the resulting rows. This yields a vector $\bar{w}$; in this case one can readily obtain an estimate for $\lambda_{\max}$ by computing $A\bar{w}$, dividing each of the components of the resulting vector by the corresponding component of $\bar{w}$, and averaging the results.

There are several useful results relating to the eigenvalue procedure. We mention a few of them here giving references where necessary.

THEOREM 1. *Let $A = (a_{ij})$ be an $n \times n$ matrix of positive coefficients with $a_{ji} = a_{ij}^{-1}$; then A is consistent if and only if $\lambda_{\max} = n$.*

Proof. From

$$\lambda = \sum_{j=1}^{n} a_{ij} w_j w_i^{-1},$$

we have

$$n\lambda - n = \sum_{\substack{i,j=1\\ i\neq j}}^{n} a_{ij} w_j w_i^{-1}.$$

It is obvious that $a_{ij} = w_i/w_j$ yields $\lambda = n$ and also $\lambda_{\max} = n$ since the sum of the eigenvalues is equal to n, the trace of A.

To prove the converse, note that in the foregoing expression we have only two terms involving a_{ij}. They are $a_{ij} w_j w_i^{-1}$ and $w_i w_j^{-1}/a_{ij}$. Their sum takes the form $y + (1/y)$.

To see that n is the minimum value of $\lambda_{\max}$ attained uniquely at $a_{ij} = w_i/w_j$ we note that for all these terms we have $y + (1/y) \geqslant 2$. Equality is uniquely obtained on putting $y = 1$, i.e., $a_{ij} = w_i/w_j$. Thus, when $\lambda_{\max} = n$ we have

$$n^2 - n \geqslant \sum_{\substack{i,j=1\\ i\neq j}}^{n} 2 = n^2 - n,$$

from which it follows that $a_{ij} = w_i/w_j$ must hold.

COROLLARY. *For a positive matrix with reciprocal entries we have*

$$\lambda_{\max} \geqslant n.$$

If A is inconsistent then we would expect that in some cases $a_{ij} \geqslant a_{kl}$ need not imply $(w_i/w_j) \geqslant (w_k/w_l)$. However, since w_i, $i = 1,..., n$, is determined by the value of an entire row, we would expect, for example, that if we have ordinal preferences among the activities, the following should hold:

THEOREM 2 (Preservation of Ordinal Consistency). *If $(o_1, ..., o_n)$ is an ordinal scale on the activities $C_1, ..., C_n$, where $o_i \geqslant o_k$ implies $a_{ij} \geqslant a_{kj}$, $j = 1,..., n$, then $o_i \geqslant o_k$ implies $w_i \geqslant w_k$.*

Proof. Indeed, we have from $Aw = \lambda_{\max} w$, that

$$\lambda_{\max} w_i = \sum_{j=1}^{n} a_{ij} w_j \geqslant \sum_{j=1}^{n} a_{kj} w_j = \lambda_{\max} w_k, \qquad \text{with} \quad w_i \geqslant w_k.$$

Because of its substantial importance, we briefly give the essential facts for the problem of existence and uniqueness of a solution to $Aw = \lambda_{\max} w$. If A is positive, the following theorem of Perron assures the existence of a solution.

THEOREM 3. *A positive matrix A has a real positive, simple "dominant" characteristic number $\lambda_{\max}$ to which corresponds a characteristic vector $w = (w_1, w_2, ..., w_n)$ of the matrix A with positive coordinates $w_i > 0$ $(i = 1, 2,..., n)$.*

When A is simply nonnegative, the theorem of Frobenius assures a similar result if A is irreducible.

DEFINITION 1. A matrix is irreducible if it cannot be decomposed into the form

$$\begin{pmatrix} A_1 & 0 \\ A_2 & A_3 \end{pmatrix}$$

where A_1 and A_2 are square matrices and 0 is the zero matrix.

The following theorem gives the equivalence of the matrix property of irreducibility and the strong connectedness of the directed graph of the matrix.

THEOREM 4. *An $n \times n$ complex matrix A is irreducible if and only if its directed graph $G(A)$ is strongly connected.*

We now state the general existence and uniqueness theorem.

THEOREM 5 (Perron–Frobenius). *Let $A \geqslant 0$ be irreducible. Then*

(i) *A has a positive eigenvalue $\lambda_{\max}$ which is not exceeded in modulus by any other eigenvalue of A.*

(ii) *The eigenvector of A corresponding to the eigenvalue $\lambda_{\max}$ has positive components and is essentially unique.*

(iii) *The number $\lambda_{\max}$ is given by.*

$$\lambda_{\max} = \max_{x \geqslant 0} \min_{1 \leqslant i \leqslant n} ((Ax)_i/x_i) = \min_{x \geqslant 0} \max_{1 \leqslant i \leqslant n} ((Ax)_i/x_i).$$

COROLLARY. *Let $A \geqslant 0$ be irreducible, and let $x \geqslant 0$ arbitrary. Then the Perron root of A satisfies*

$$\min_{1 \leqslant i \leqslant n} ((Ax)_i/x_i) \leqslant \lambda_{\max} \leqslant \max_{1 \leqslant i \leqslant n} ((Ax)_i/x_i).$$

A well-known theorem of Wielandt (1950) in matrix theory yields a stronger result than the following, which may be taken as a corollary to it:

If A is a nonnegative irrducible matrix, then the value of $\lambda_{\max}$ increases with any element a_{ij} of A.

This corollary does not say explicitly how $\lambda_{\max}$ increases with a_{ij}. However, an interesting observation for our purpose is that while an increase in a_{ij} gives rise to an increase in $\lambda_{\max}$, this increase is partly offset by a decrease in $a_{ji} = 1/a_{ij}$ which is one of the requirements in filling out the comparison matrix A.

It is known that the normalized row of the limiting matrix of A^k corresponds to the normalized eigenvector of $Aw = \lambda_{\max} w$. There are several ways of proving this. The simpler proofs require special assumptions on the eigenvectors of A.

DEFINITION 2. We define the norm of the matrix A by $\| A \| \equiv (A\bar{e})^T \bar{e}$ (i.e., it is the sum of all entries of A), where

$$\bar{e} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}.$$

DEFINITION 3. A nonnegative irreducible matrix A is primitive if and only if there is an integer $p \geqslant 1$ such that $A^p > 0$.

THEOREM 6. *For a primitive matrix A*

$$\lim_{k\to\infty}(A^k\bar{e}/\| A^k \|) = Cw_{\max},$$

where C is a constant and $w_{\max}$ is the normalized eigenvector corresponding to $\lambda_{\max}$.

The following theorem asserts that the ratios of normalized eigenvector components remain the same when any row and corresponding column are deleted from a consistent matrix of pairwise comparisons.

THEOREM 7. *If A is a positive consistent matrix and A' is obtained from A by deleting the ith row and ith column then A is consistent and its corresponding eigenvector is obtained from that of A by putting $w_i = 0$ and normalizing the components.*

Proof. Given any row of A, e.g., the first, we have $a_{ij} = a_{1j}/a_{1i}$, $j = 1,..., n$. Thus the ith row of A depends on the ith column entry in its first row being given. Conversely, $a_{jk} = a_{1k}/a_{1j}$. Thus no entry in A' depends on the ith row or ith column of A and hence A' is also consistent. Since their entries coincide except in the ith row and ith column of A and since the solution of an eigenvalue problem with a consistent matrix is obtained from any normalized column, the theorem follows.

Remark. In the general case, if $A = (a_{ij})$ is a matrix of pairwise comparisons and $A' = (a'_{ij})$ with $a'_{ij} = a_{ij}$, $i, j = 1,..., n$, $i \neq k$, $j \neq k$, $a'_{ij} = 0$, $i = k$ or $j = k$, and if the normalized eigenvector solutions of $Aw = \lambda_{\max}w$ and $A'w' = \lambda_{\max}w'$ are w and w', respectively, then $w_k' = 0$ but $w_\alpha'/w_\beta' \neq w_\alpha/w_\beta$, for all α and β. In other words leaving one activity out of a pairwise comparison matrix does not distribute its weight proportionately among the other activities. The reason can be seen from the limiting relations which show that each activity is involved with the others in a complicated way.

EXAMPLE. Here we are only interested in numerical entries of the eigenvector. In measurement of the relative wealth of nations illustration given later, the USSR, which occupies the second entry, is in the first comparison but is taken out in the second, retaining the others. No proportionality equivalence is observed.

U.S.	USSR	China	France	U.K.	Japan	W. Germany
0.427	0.230	0.021	0.052	0.052	0.123	0.094
0.504	0.0	0.0258	0.0728	0.0728	0.184	0.140

The following theorem shows that seeking an order type of relationship between a_{ij} and w_i/w_j involves all of A and its powers in a complicated fashion.

THEOREM 8. *For a primitive matrix A we have $a_{ij} \geqslant a_{kl}$ if and only if $w_i/w_j \geqslant w_k/w_l$, whenever*

$$\lim_{m\to\infty}\Big(\sum_{p\neq j} a_{ip}(A^m e)_p/(A^m e)_j\Big) \geqslant \lim_{m\to\infty}\Big(\sum_{q\neq l} a_{kq}(A^m e)_q/(A^m e)_l\Big)$$

holds. (A pth subscript on a vector indicates the use of its pth entry.)

Proof. In a typical case

$$\sum_{j=1}^{n} a_{ij}w_j = \lambda w_i ,$$

from which we have

$$a_{ij} = \frac{\lambda w_i}{w_j} - \frac{1}{w_j}\sum_{p\neq j} a_{ip}w_p ,$$

It also follows that

$$a_{kl} = \frac{\lambda w_k}{w_l} - \frac{1}{w_l}\sum_{q\neq l} a_{kq}w_q ,$$

$$a_{ij} \geqslant a_{kl} \leftarrow \frac{\lambda w_i}{w_j} \geqslant \frac{\lambda w_k}{w_l} + \frac{1}{w_j}\sum_{p\neq j} a_{ip}w_p - \frac{1}{w_l}\sum_{q\neq l} a_{kq}w_q .$$

Thus, the theorem is true whenever the following inequality holds:

$$(1/w_j)\sum_{p\neq j} a_{ip}w_p \geqslant (1/w_l)\sum_{q\neq l} a_{kq}w_q .$$

Using Theorem 6 we replace every w_s by

$$\lim_{m\to\infty}((A^m e)_s/e^T A^m e),$$

yielding the proof.

Assume that our mind in fact works with pairwise comparisons but the a_{ij} are not estimates of w_i/w_j but of some function of the latter, $a_{ij}(w_i/w_j)$. For example, Stevens (1959) observed that a_{ij} as perceived for prothetic phenomena takes the form $(w_i/w_j)^a$, where a lies somewhere between 0.3 (in the case of loudness estimation) and 4 (in the case of electric shock estimation). For metathetic phenomena, Stevens points out that the power law need not apply, i.e., $a = 1$.

Thus it is of interest to study the general form of the solution $g_i(w_i)$, $i = 1,..., n$, of an eigenvalue problem satisfying the generalized consistency condition of the form

$$f(a_{ij})\, f(a_{jk}) = f(a_{ik}).$$

THEOREM 9 (The Eigenvalue Power Law). *If the matrix $A = (a_{ij}(w_i/w_j))$ of order n satisfies the generalized consistency condition, then the eigenvalue problem*

$$\sum_{j=1}^{n} a_{ij}(w_i/w_j)\, g_j(w_j) = n g_i(w_i), \qquad i = 1,..., n,$$

has the eigenvector solution $(w_1{}^a,..., w_n{}^a) \equiv (g_1(w_1),..., g_n(w_n))$.

Proof. Substituting the relation

$$a_{ij}(w_i/w_j) = g_i(w_i)/g_j(w_j)$$

satisfied by the solution $g_i(w_i)$, $i = 1,..., n$, of the eigenvalue problem into the consistency condition we have

$$f[g_i(w_i)/g_j(w_j)]\, f[g_j(w_j)/g_k(w_k)] = f[g_i(w_i)/g_k(w_k) \cdot g_j(w_j)/g_j(w_j)],$$

or if we put

$$x = g_i(w_i)/g_j(w_j) \qquad \text{and} \qquad y = g_j(w_j)/g_k(w_k),$$

we have

$$f(x)\, f(y) = f(xy).$$

This functional equation has the general solution

$$f(x) = x^a.$$

Thus generalizing the consistency condition for A implies that a generalization of the corresponding eigenvalue problem (with $\lambda_{\max} = n$) is solvable if we replace a_{ij} by a constant power a of its argument. But we know that when $a = 1$, $a_{ij} = w_i/w_j$; thus, in general, $a_{ij} = (w_i/w_j)^a$, which implies that

$$g_i(w_i)/g_j(w_j) = (w_i/w_j)^a, \qquad i, j = 1,..., n,$$

and hence,

$$g_i(w_i) = w_i{}^a = g(w_i), \qquad i = 1,..., n.$$

Thus the solution of a pairwise comparison eigenvalue problem satisfying consistency produces estimates of a power of the underlying scale rather than the scale itself. In applications where knowledge, rather than our senses, is used to obtain the data, one would expect the power to be equal to unity and, hence, we have an estimate of the underlying scale itself. This observation may be useful in social applications.

3. The Scale

We now discuss the scale we recommend for use which has been successfully tested and compared with other scales.

The judgments elicited from prople are taken qualitatively and corresponding scale values are assigned to them. In general, we do not expect "cardinal" consistency to hold everywhere in the matrix because people's feelings do not conform to an exact formula. Nor do we expect "ordinal" consistency, as people's judgments may not be transitive. However, to improve consistency in the numerical judgments, whatever value a_{ij} is assigned in comparing the ith activity with the jth, the reciprocal value is assigned to

a_{ji}. Thus we put $a_{ji} = 1/a_{ij}$. Usually we first record whichever value represents dominance greater than unity. Roughly speaking, if one activity is judged to be α times stronger than another, then we record the latter as only $1/\alpha$ times as strong as the former. It can be easily seen that when we have consistency, the matrix has unit rank and it is sufficient to know one row of the matrix to construct the remaining entries. For example, if we know the first row then $a_{ij} = a_{1j}/a_{1i}$ (under the rational assumption of course, that $a_{1i} \neq 0$ for all i).

It is useful to repeat that reported judgments need not be even ordinally consistent and, hence, they need not be transitive; i.e., if the relative importance of C_1 is greater than that of C_2 and the relative importance of C_2 is greater than that of C_3, then the relation of importance of C_1 need not be greater than that of C_3, a common occurrence in human judgments. An interesting illustration is afforded by tournaments regarding inconsistency or lack of transitivity of preferences. A team C_1 may lose against another team C_2 which has lost to a third team C_3; yet C_1 may have won against C_3. Thus, team behavior is inconsistent—a fact which has to be accepted in the formulation, and nothing can be done about it.

We now turn to a question of what numerical scale to use in the pairwise comparison matrices. Whatever problem we deal with we must use numbers that are sensible. From these the eigenvalue process would provide a scale. As we said earlier, the best argument in favor of a scale is if it can be used to reproduce results already known in physics, economics, or in whatever area there is already a scale. The scale we propose is useful for small values of $n < 10$.

Our choice of scale hinges on the following observation. Roughly, the scale should satisfy the requirements:

1. It should be possible to represent people's differences in feelings when they make comparisons. It should represent as much as possible all distinct shades of feeling that people have.

2. If we denote the scale values by $x_1, x_2, \ldots, x_p$, then let

$$x_{i+1} - x_i = 1, \qquad i = 1, \ldots, p - 1.$$

Since we require that the subject must be aware of all gradations at the same time, and we agree with the psychological experiments (Miller, 1956) which show that an individual cannot simultaneously compare more than seven objects (plus or minus two) without being confused, we are led to choose a $p = 7 + 2$. Using a unit difference between successive scale values is all that we allow, and using the fact that $x_1 = 1$ for the identity comparison, it follows that the scale values will range from 1 to 9.

As a preliminary step toward the construction of an intensity scale of importance for activities, we have broken down the importance ranks as shown in the following scale (Table 1). In using this scale the reader should recall that we assume that the individual providing the judgment has knowledge about the relative values of the elements being compared whose ratio is $\geqslant 1$, and that the numerical ratios he forms are nearest-integer approximations scaled in such a way that the highest ratio corresponds to 9. We have assumed that an element with weight zero is eliminated from comparison. This, of course,

TABLE 1

The Scale and Its Desceiption

Intensity of importance	Definition	Explanation
1[a]	Equal importance	Two activities contribute equally to the objective
3	Weak importance of one over another	Experience and judgment slightly favor one activity over another
5	Essential or strong importance	Experience and judgment strongly favor one activity over another
7	Demonstrated importance	An activity is strongly favored and its dominance is demonstrated in practice.
9	Absolute importance	The evidence favoring one activity over another is of the highest possible order of affirmation
2, 4, 6, 8	Intermediate values between the two adjacent judgments	When compromise is needed
Reciprocals of above nonzero	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	
Rationals	Ratios arising from the scale	If consistency were to be forced by obtaining n numerical values to span the matrix

[a] On occassion in 2 by 2 problems, we have used $1 + \epsilon$, $0 < \epsilon \leqslant \frac{1}{2}$ to indicate very slight dominance between two nearly equal activities.

does imply that zero may not be used for pairwise comparison. Reciprocals of all scaled ratios that are $\geqslant 1$ are entered in the transpose positions (not taken as judgments). Note that the eigenvector solution of the problem remains the same if we multiply the unit entries on the main diagonal, for example, by a constant greater than 1.

At first glance one would like to have a scale extend as far out as possible. On second thought we discover that to give an idea of how large measurement can get, scales must be finite. We also note that one does not measure widely disparate objects by the same yardstick. Short distances on a piece of paper are measured in centimeters, longer distances in a neighborhood in meters, and still larger ones in kilometers and even in light years. To make comparisons of the sizes of atoms with those of stars, people, in a natural fashion, insert between thse extremes, objects which gradually grow larger and larger enabling one to appreciate the transition in the magnitudes of measurement. To make such a transition possible the objects are divided into groups or clusters whereby the objects put into each group are within the range of the scale and the largest object in one group is used as the smallest one in the next larger group. Its scale values in the two groups enable one to continue the measurement from one group to the next and so on. We have more to say about clustering in a later section.

In practice, one way or another, the numerical judgments have to be approximations, but how good these approximations are is the question to which our theory is aimed.

A typical question to ask in order to fill in the entries in a matrix of comparisons is: Consider two activities i on the left side of the matrix and another j on the top; which of the two has the property under discussion more, and how strongly more (using the scale values 1 to 9)? This gives us a_{ij}. The reciprocal value is then automatically entered for a_{ji}.

It should be noted that consistency is a necessary but not a sufficient condition for judging how good a set of observational data is. The consistency may be good, but the correspondence of the judgments to reality may be poor. We have already discussed how one may decide on the goodness of consistency by comparing the mean μ with the standard deviation $(2\mu)^{1/2}$ and requiring that their ratio be $\leqslant 1$. Alternatively, one may accept H_0 on comparing μ with its average value from the solutions of a sizable sample of eigenvalue problems whose matrices have random entries from the scale 1 to 9 using the reciprocal value in the transpose position. For these averages for matrices of different order see the discussion below.

As yet there is no statistical theory (to the best of our knowledge) which would assist us in deciding how well judgmental data correspond to reality. We have occasionally used the root mean square deviation (RMS) and the median absolute deviation about the median (MAD). These indicators are probably more useful in making interscale or interpersonal comparisons in judgments than as absolute measures of the goodness of fit. We have not found the χ^2 test useful. It is clear that this is an area of research that is worth pursuing.

Considerable effort has been concentrated on comparing the scale 1 to 9 with 25 other scales suggested to us by a number of people. We took pairwise qualitative judgments described in our scale including qualities between those mentioned in the table. For example, a judge could simply say in a comparison that it is between equal and weak or between weak and strong, etc. Five applications were made. Three of them were examples described below: distance estimation, optics, and wealth. We replaced the qualities by numbers from each scale uniformly distributed over the qualities, sometimes leaving gaps as many of the scales had a wide range. For each scale we then calculated the eigenvector after using the reciprocal property in the matrix. Our object was to compare all these eigenvectors for each example with the true answer, which was known, and compare the RMS and MAD. On the whole we had the best results for the scale 1 to 9 even though the consistency for this scale was not always the best.

It also turns out that a good judge gives good results by any scale including direct estimation. A judge who is not an expert can see in the pairwise comparison process where his judgment is strong and where it is weak. The eigenvalue approach is excellent for bargaining purposes as it permits people to debate the reasons for their estimates, arrive at a consensus, and make compromises here and there. In over 30 applications of the process, we noted that people were generally very content with the interaction and the outcome.

Figure 1a and its associated table (Table 2) are of interest. For each order matrix we constructed a sample of size 50 and filled in its entries at random from the scales 1–5, 1–7, 1–9, 1–15, 1–20, and 1–90. Thus, for example, for the scale 1–5, the main diagonal entries are unity and for each position above the diagonal we chose any of the integers 1–5 or their reciprocals at random. The reciprocal of this entry was then given to its transpose. The same procedure was carried out for the other scales. We averaged $(\lambda_{\max} - n)/(n - 1)$ for the 50 matrices corresponding to each value of n and for each scale.

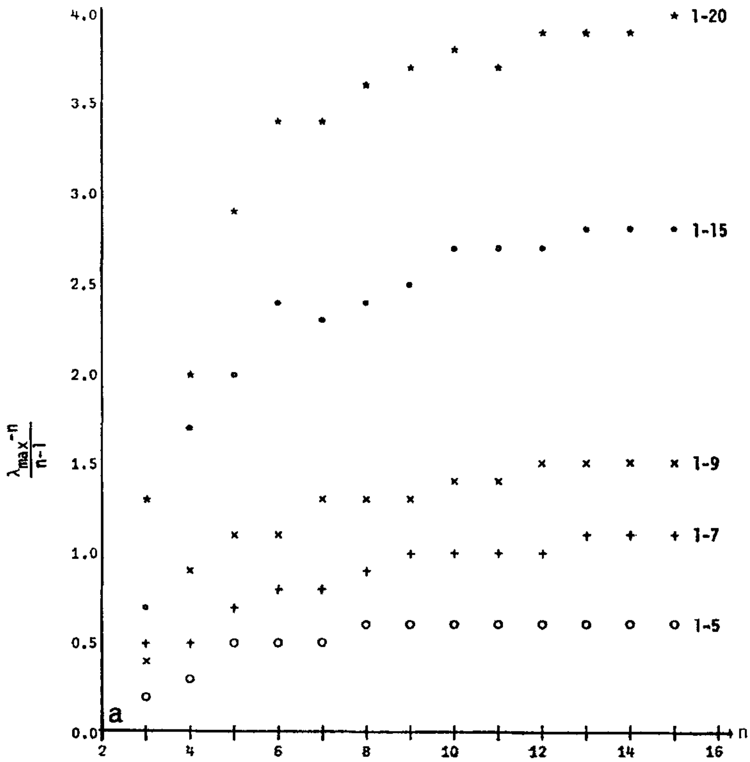

FIG. 1a. Average consistency for matrices with random entries.

We obtained the Table 2 and Fig. 1a, which corresponds to it. This table is useful for comparing the significance of the inconsistency calculated for a particular problem with the average value obtained for the scale being used. In our case the relevant values are for the scale 1–9. In this comparison we can require the ratio to be very small; e.g., of the order of 0.1. Note that a plot of the row in the table corresponding to the scale 1–90 is not given as the original drawings carried the plot beyond the page.

We now make another interesting observation using this result. It is generally known that if λ is any of the eigenvalues of a matrix, then $| \lambda - a_{ii} | \leqslant \sum_{j \neq i} | a_{ij} |$ for some i, $i = 1,..., n$.

TABLE 2

Measure of Inconsistency μ

Scale	Order of matrix													
	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1–5	0.000	0.244	0.335	0.472	0.479	0.527	0.580	0.577	0.611	0.591	0.623	0.632	0.641	0.629
1–7	0.000	0.515	0.504	0.708	0.798	0.827	0.922	0.961	0.968	1.012	1.019	1.054	1.052	1.052
1–9	0.000	0.416	0.851	1.115	1.150	1.345	1.334	1.315	1.420	1.395	1.482	1.491	1.470	1.466
1–15	0.000	0.705	1.733	2.024	2.416	2.349	2.351	2.525	2.674	2.749	2.693	2.804	2.827	2.806
1–20	0.000	1.326	2.044	2.948	3.354	3.428	3.598	3.709	3.807	3.719	3.899	3.888	3.895	3.971
1–90	0.000	3.206	10.411	15.452	16.096	17.603	17.454	18.580	19.110	18.747	19.695	19.857	19.990	20.052

Since for a reciprocal positive matrix $\lambda_{\max} \geqslant n$ and $a_{ii} = 1$, we may simply write

$$\lambda_{\max} \leqslant \max_i \sum_{j=1}^{n} a_{ij} .$$

Now the maximum value for any a_{ij} when we use the scale 1–9 is 9. Thus $\lambda_{\max}$ is at most equal to $9n$. We also note that $(\lambda_{\max} - n)/(n-1) \leqslant 8(n/(n-1))$ and is therefore bounded above. In fact, as $n \to \infty$ one can show that $\mu = (\lambda_{\max} - n)/(n-1)$ tends to a limit, a result confirmed by our statistical approach. We leave the theoretical calculation of this asymptotic limit for a separate paper. What we have done (instead of using difference methods) is to take the average of the last three values, i.e., for $n = 13, 14, 15$ in Table 1 for each scale, and use it as an approximation to the limiting value. If we denote this value by L_s for scale s, we then calculate a new table using $C = (L_s - \mu)/L_s$ for each n. C measures consistency expressed as an index between zero and unity. This leads to Table 3 and its associated graph, Fig. 1b.

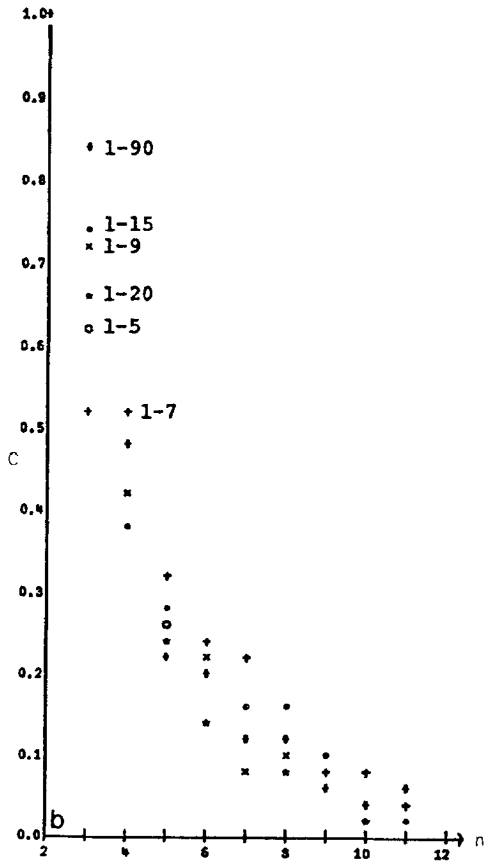

FIG. 1b. Consistency normalized using asymptotic value.

TABLE 3

Normalized Measure of Consistency C

Scale	Order of matrix									
	2	3	4	5	6	7	8	9	10	11
1–5	1.000	0.616	0.472	0.255	0.245	0.169	0.035	0.089	0.036	0.068
1–7	1.000	0.511	0.522	0.328	0.242	0.214	0.124	0.087	0.080	0.039
1–9	1.000	0.718	0.424	0.244	0.221	0.088	0.096	0.109	0.037	0.055
1–15	1.000	0.749	0.384	0.280	0.141	0.165	0.160	0.102	0.049	0.023
1–20	1.000	0.662	0.478	0.248	0.144	0.125	0.082	0.053	0.028	0.051
1–90	1.000	0.839	0.479	0.226	0.194	0.118	0.126	0.069	0.043	0.061

Now this is the consistency measured for randomly filled matrices. In general, informed judgment leads to better consistency. However, all the plots show that when the number of objects being compared exceeds 7 ± 2, the consistency can be expected to be very poor—a theoretical confirmation of Miller's psychological observation. Later on we show how to overcome this limitation on the number of objects by using a method of hierarchical clustering.

We have extended the method to the construction of a matrix of data from an incomplete set of judgments. In some cases, regression analysis or row and column means have been used to estimate the missing entries, but this differs from the way we approach the problem. Our use of hierarchical decomposition allows one to group N objects into a manageable number of clusters, perform the pairwise comparisons on these clusters, obtaining an eigenvector of weights for them. One then decomposes each cluster into smaller clusters and so on, each time performing similar operations, but then composing the results in a manner to be discussed later to obtain an overall weighting. Thus, in general, the number of comparisons required is considerably less than $(N^2 - N)/2$. However, we can also simplify the task in each matrix by asking for judgments to be supplied in n positions forming a spanning cycle of the complete graph of our matrix. The n judgments thus obtained may be used to construct the entire matrix through the relation $a_{ij}a_{jk} = a_{ik}$ which plays an important role in our theory. Our consistency index is then calculated. If it is small, the corresponding eigenvector is the desired solution. Otherwise, a matrix of ratios of eigenvector values is constructed, absolute differences from the original matrix are computed, and a new judgment is obtained for the largest difference entry. The iterations are continued using the new matrix with a view to improving consistency. Thus, pursuit of consistency remains central in our approach to the incomplete data problem. The procedure requires testing, as we have often filled the entire set of $(n^2 - n)/2$ values in each cluster.

Examples

1. *Distance Estimation through Air Travel Experience*

The Eigenvalue method was used to estimate the relative distances of six cities from Philadelphia by making pairwise comparisons between them as to which was how strongly farther from Philadelphia.

It is interesting to note that the cities cluster into three classes—those nearest to Philadelphia: Montreal and Chicago; those which are intermediate: San Francisco and London; and those farthest: Cairo and Tokyo. The latter, because of relatively large value due to errors of uncertainty, cause the values of the others to be perturbed from where we want them to be. Thus, if their eigenvector components change comparatively slightly, and the increment is distributed among the others, the relative values of these can be altered considerably.

The next matrix (Table 4) gives numerical values to the perceived remoteness from Philadelphia for each pair of cities. The rows indicate the strength of dominance. The question to ask is: Given a city (on the left) and another (on top) how much strongly farther is the first one from Philadelphia than the second? We then put the reciprocal value in the transpose position. Compare the solution of the eigenvalue problem with the actual result given in Table 5. (For a more detailed explanation of the use of the scale in an application see the explanation under the matrix of Example 3 below.) We have $\lambda_{\max} = 6.45$. Our qualitative statistical test shows that the consistency is good since $(\mu/2)^{1/2} = (0.09/2)^{1/2} = 0.21 < 1$. Also, comparison of inconsistency with the value for $n = 6$ (scale 1–9) from Table 2 gives $0.09/1.15 = 0.08$, which is good.

TABLE 4

Comparisons of Distances of Cities from Philadelphia

	Cairo	Tokyo	Chicago	San Francisco	London	Montreal
Cairo	1	$\frac{1}{3}$	8	3	3	7
Tokyo	3	1	9	3	3	9
Chicago	$\frac{1}{8}$	$\frac{1}{9}$	1	$\frac{1}{6}$	$\frac{1}{5}$	2
San Francisco	$\frac{1}{3}$	$\frac{1}{3}$	6	1	$\frac{1}{3}$	6
London	$\frac{1}{3}$	$\frac{1}{3}$	5	3	1	6
Montreal	$\frac{1}{7}$	$\frac{1}{9}$	$\frac{1}{2}$	$\frac{1}{6}$	$\frac{1}{6}$	1

Besides the largest eigenvalue 6.45, this example has the following remaining eigenvalues: -0.260, $-0.230 + 0.665i$, $-0.230 - 0.665i$, $0.133 + 1.577i$, $0.133 - 1.577i$.

2. *Illumination Intensity and the Inverse Square Law*

The rate at which a source emits light energy evaluated in terms of its visual effects is spoken of as light flux. The illumination of a surface is defined as the amount of light flux it receives per unit area.

TABLE 5

Actual and Judgmental Distances

City	Distance to Philadelphia (miles)	Normalized distance	Eigenvector
Cairo	5729	0.278	0.263
Tokyo	7449	0.361	0.397
Chicago	660	0.032	0.033
San Francisco	2732	0.132	0.116
London	3658	0.177	0.164
Montreal	400	0.019	0.027

The following experiment was conducted in search of a relationship between the illumination received by four identical objects (placed on a line at known distances from a light source) and of the distance from the source. The comparison of illumination intensity was performed visually and independently by two sets of people. The objects were placed at the following distances measured in yards from the light source: 9, 15, 21, and 28. In normalized form these distances are: 0.123, 0.205, 0.288, 0.384.

The two matrices of pairwise comparisons of the brightness of the objects labeled in increasing order according to their nearness to the source where the judges were located are:

Relative visual brightness (1st trial)

	C_1	C_2	C_3	C_4
C_1	1	5	6	7
C_2	$\frac{1}{5}$	1	4	6
C_3	$\frac{1}{6}$	$\frac{1}{4}$	1	4
C_4	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{4}$	1

Relative visual brightness (2nd trial)

	C_1	C_2	C_3	C_4
C_1	1	4	6	7
C_2	$\frac{1}{4}$	1	3	4
C_3	$\frac{1}{6}$	$\frac{1}{3}$	1	2
C_4	$\frac{1}{7}$	$\frac{1}{4}$	$\frac{1}{2}$	1

Relative brightness eigenvector 1st trial ($\lambda_{\max} = 4.39$)	Relative brightness eigenvector 2nd trial ($\lambda_{\max} = 4.1$)
0.61	0.62
0.24	0.22
0.70	0.10
0.05	0.06

The top two columns of Table 6 should be compared with the bottom-right column calculated from the inverse square law in optics.

TABLE 6

Inverse Square Law of Optics

Distance	Normalized distance	Square of normalized distance	Recriprocal of square column	Normalized recriprocal	Rounding off
9	0.123	0.015129	66.098	0.6079	0.61
15	0.205	0.042025	23.79	0.2188	0.22
21	0.288	0.082944	12.05	0.1108	0.11
28	0.384	0.147456	6.78	0.0623	0.06

Note the sensitivity of the results as the object is very close to the source for then it absorbs most of the value of the relative index and a small error in its distance from the source yields great error in the values. What is noteworthy from this sensory experiment is the observation or hypothesis that the observed intensity of illumination varies (approximately) inversely with the square of the distance. The more carefully designed the experiment, the better the results obtained from the visual observation.

It may be of interest to mention in passing that the eigenvalues of the second pairwise comparison matrix above besides the largest, which is 4.1, are -0.078, $-0.012 + 0.646i$ and its complex conjugate is $-0.012 - 0.646i$.

This experiment was repeated with more objects and different distances, again with good results. Some psychophysicist colleagues have felt that if great distances had been used in the experiment the result would not have been as good because Stevens' power law has exponent $\neq 1$ for brightness. We feel that long distances and many objects could be composed into clusters hierarchically and our procedure, described below, might still apply. Perhaps people "compensated" for distance in judging "illumination." If $s = I^\alpha$ is the "true" brightness sensation, $J = D^\beta s$ is the "compensation formula," and $I = D^{-2}$ is the inverse square law, then $\beta - 2^\alpha = -2$ if $J = D^{-2}$; $\alpha = 0.25$ to 0.5 in Stevens' work: If $\beta = -1$, $\alpha = 0.5$. If $\beta = -1.5$, $\alpha = 0.25$.

3. *The Wealth of Nations through Their World Influence*

A number of people have studied the problem of measuring the world influence of nations. We have briefly examined this concept within the framework of our model. We assumed that influence is a function of several factors. We considered (Saaty and Khouja, 1976) five such factors: (1) human resources; (2) wealth; (3) trade; (4) technology; and (5) military power. Culture and idealogy, as well as potential natural resources (such as oil), were not included.

Seven countries were selected for this analysis. They are the United States, the U.S.S.R., China, France, the United Kingdom, Japan, and West Germany. It was felt that these nations as a group comprised a dominant class of influential nations. It was desired to compare them among themselves as to their overall influence in international relations. We realize that what we have is a very rough estimate—mainly intended to serve as an interesting example of an application of our approach to priorities. We only illustrate the method with respect to the single factor of wealth. The question to answer is: how much more strongly does one nation as compared with another contribute its wealth to gain world influence?

TABLE 7

Wealth Comparison of Nations

	U. S.	USSR	China	France	U. K.	Japan	W. Germany
U. S.	1	4	9	6	6	5	5
USSR	$\frac{1}{4}$	1	7	5	5	3	4
China	$\frac{1}{9}$	$\frac{1}{7}$	1	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{7}$	u
France	$\frac{1}{6}$	$\frac{1}{5}$	5	1	1	$\frac{1}{3}$	$\frac{1}{3}$
U. K.	$\frac{1}{6}$	$\frac{1}{5}$	5	1	1	$\frac{1}{3}$	$\frac{1}{3}$
Japan	$\frac{1}{5}$	$\frac{1}{3}$	7	3	3	1	2
W. Germany	$\frac{1}{5}$	$\frac{1}{4}$	5	3	3	$\frac{1}{2}$	1

The first row of Table 7 gives the pairwise comparison of the wealth contributed by the United States with the other nations. For example, it is of equal importance to the United States (hence, the unit entry in the first position), between weak and strong importance when compared with the U.S.S.R. (hence, the value 4 in the second position), of absolute importance when compared with China (hence, the value 9 in the third position). We have values between strong and demonstrated importance when compared with France and the United Kingdom (hence, a 6 in the next two positions), strong importance when compared with Japan and Germany (hence a 5 in the following two positions). For the entries in the first column we have the reciprocals of the numbers in the first row indicating the inverse relation of relative strength of the wealth of the other countries when compared with the United States, and so on, for the remaining values in the second row and second column, etc.

Note that the comparisons are not consistent. For example, U.S. : U.S.S.R. = 4, U.S.S.R. : China = 7, but U.S. : China = 9, not 28.

Nevertheless, when the requisite computations are performed, we obtain relative weights of 42.9 and 23.1 for the United States and Russia, and these weights are in striking agreement with the corresponding GNP's as percentages of the total GNP (see Table 8). Thus, despite the apparent arbitrariness of the scale, the irregularities disappear and the numbers occur in good accord with observed data.

The largest eigenvalue of the wealth example is 7.61 and the remaining eigenvalues are -0.228, 2×10^{-11}, $-0.330 + 0.588i$, $-0.330 - 0.588i$, $0.14 + 2.06i$, $0.14 - 2.06i$.

TABLE 8

Normalized Wealth Eigenvector

	Normalized eigenvector	Actual GNP[a] (1972)	Fraction of GNP Total
U. S.	0.429	1167	0.413
USSR	0.231	635	0.225
China	0.021	120	0.043
France	0.053	196	0.069
U. K.	0.053	154	0.055
Japan	0.119	294	0.104
W. Germany	0.095	257	0.091
Total		2823	

[a] Billions of dollars.

Compare the normalized eigenvector column derived by using the matrix of judgments in Table 8 with the actual GNP fraction given in the last column. The two are very close in their values. Estimates of the actual GNP of China range from 74 billion to 128 billion. Cluster analysis can be used to show that China probably should not be in the group.

4. *Weight Estimation*

First, we compared five objects in pairs by picking them up one at a time to get an idea of the range of their weight intensities; then we compared all of the objects with each one by picking them up with the right hand one at a time. The objects and the matrix of pairwise comparison are as shown in Table 9.

TABLE 9

Weight Comparison of Objects

	Radio	Typewriter	Large attache case	Projector	Small attache case
Radio	1	$\frac{1}{5}$	$\frac{1}{3}$	$\frac{1}{4}$	4
Typewritter	5	1	2	2	8
Large attache case	3	$\frac{1}{2}$	1	$\frac{1}{2}$	4
Projector	4	$\frac{1}{2}$	2	1	7
Small attache case	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{7}$	1

The resulting eigenvector and vector of actual relative weight are given by

Eigenvector	Actual
0.09	0.10
0.40	0.39
0.18	0.20
0.29	0.27
0.04	0.04

The root mean square deviation is given by 0.0158, less than a 2% error, indicating a good estimate. It is worth noting that comparing objects to estimate their weights by lifting them up with the hand is something people rarely do. Thus, one would expect wider scatter in the results than, for example, in an optics experiment where the eyes are used to compare the relative brightnesses of objects from a light source, something the eye does all the time. Greater precision is expected from the eye because of its experience.

The largest eigenvalue is 5.16 and the remaining ones are $-0.056 + 0.433i$, $-0.056 - 0.433i$, $-0.025 + 0.798i$, $-0.025 - 0.798i$.

4. Hierarchies—General Considerations

Although most people have an idea of what a hierarchy is, few use the concept in their thinking. Fewer still realize how important and powerful a hierarchy is as a model of reality when viewing a complex system of interacting components.

First, we give an overview of the idea of a hierarchy, its application to real systems and to thought processes, and its usefulness as a general model. We then present a formal theory for the analysis of hierarchies and their stability and explain what to do with the model after it is obtained. Although the latter research is relatively new, it has been put to use in a number of practical applications.

Among these we mention: (1) a development of a theory of priorities, (2) a theory of two-point boundary planning (forward and backward processes), and (3) a new method in conflict resolution.

Any system can be represented by a large interaction matrix whose rows and columns are components of the system. When component i and component j interact strongly, the i, jth entry is near ± 1. When they do not interact, the entry is near zero. In a large system, most of the entries are close to zero. Using the concept of a reachability matrix and its powers, a distinct hierarchic structure is often discerned. In fact, this arrangement of the elements of a system in an incidence type of matrix can be used to identify the levels of a hierarchy. We do not describe this well-known process here because it usually produces results in line with one's intuition about what falls in which levels. In the simplest type of hierarchy an upper level dominates the neighboring lower level.

Different levels of a hierarchy are generally characterized by differences in both structure and function. The proper functioning of a higher level depends on the proper functioning of the lower levels. The basic problem with a hierarchy is to seek understanding at the highest levels from interactions of the various levels of the hierarchy rather than directly from the elements of the levels. Rigorous methods for structuring systems into hierarchies are gradually emerging in the natural and social sciences and in particular, in general systems theory as it relates to the planning and design of social systems.

Hierarchies are order-preserving structures. They involve the study of order among partitions of a set. The partitions are called the levels of the hierarchy. Conceptually the simplest hierarchy is linear, rising from one level to an adjacent level. The complexity of the arrangement of the elements in each level may be the same or it may increase from level to level. This also applies to the depth of analytical detail. A hierarchy may emerge gradually from one root (e.g., the development of the human race from a first man) or it may descend in rank from one boss, as in an organization. It may grow by adding parts like a snowball or it may be a simple gradual arrangement of the levels according to a pattern. The structure of each level may take the form of a general network representing the appropriate connections among its elements. This last subject is of considerable interest to us in developing a methematical theory of hierarchies as we obtain a method for evaluating the impact of a level on an adjacent level from the interactions of the elements in that level.

Perhaps one of the most interesting analytical works which have enriched the concept of hierarchy is the paper by Simon and Ando (1961), from which the first author derived many insights into hierarchies for his subsequent works on the subject.

A simple comprehensive example of a hierarchy begins with the entire universe as one level, galactic clusters as the next level, then successively to galaxies, constellations, solar systems, planets, clumps of matter, crystals, compounds, molecular chains, molecules, atoms, nucleii, protons, and neutrons. Another example of a hierarchy is that representing the structure of living organism, and a third example would be one which represents the functions of an organizational hierarchy. The following are two illustrations. (See Fig. 2 and 3.) A hierarchy is complete when each level connects to all elements in the next higher level.

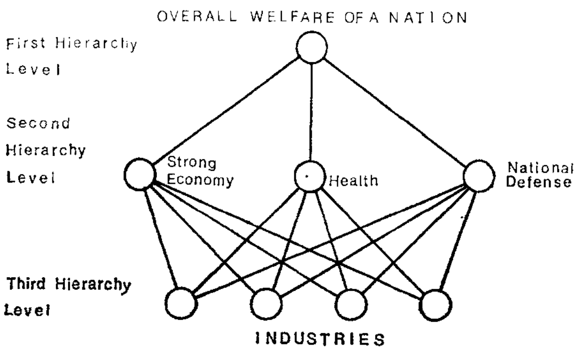

Fig. 2. A complete hierarchy for priorities of industries.

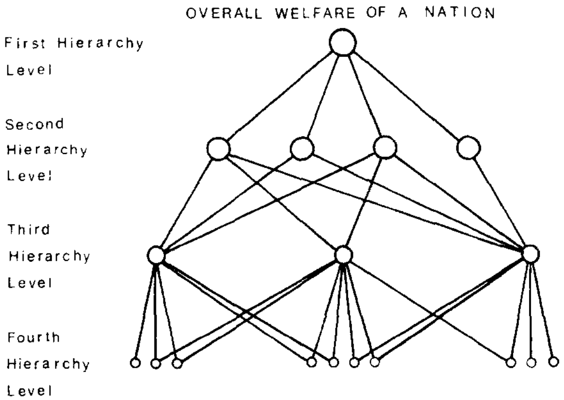

FIG. 3. A hierarchy for priorities of transport projects in national planning.

In Fig. 2 the first hierarchy level has a single objective, the overall welfare of a nation. Its priority value is assumed to be equal to unity. The second hierarchy level has three objectives, strong economy, health, and national defense. Their priorities are derived from a matrix of pairwise comparisons with respect to the objective of the first level. The third hierarchy level objectives are the industries. The object is to determine the impact of the industries on the overall welfare of a nation through the intermediate second level. Thus their priorities with respect to each objective in the second level are obtained from a pairwise comparison matrix with respect to that objective and the resulting three priority vectors are then weighted by the priority vector of the second level to obtain the desired composite vector of priorities of the industries.

In Fig. 3 the hierarchy consists of four levels: the first is the overall welfare of a nation; the second, a set of possible future scenarios of that nation; the third, the provinces of that nation; and the fourth, transport projects which fall in the provinces. Note that here not every province affects each future scenario nor does each project affect every province. The hierarchy in Fig. 3 is not a complete one. The object here is to determine the priorities of the projects as they impact on the overall objective. Here one must weight the priorities of each comparison set by the ratio of the number of elements in that set to the total number of elements in the fourth level. This is what one has to do when the hierarchy is not complete.

Hierarchies may be used to represent both the *structural* and *functional relations* of a system. Because of the close identification of hierarchies; i.e., autonomous, dynamic, etc., each element of a given hierarchy may belong functionally to several other different hierarchies. A spoon may be arranged with order spoons of different sizes in one hierarchy or with knives and forks in a second hierarchy. For example, it may be a controlling component in a level of one hierarchy or it may simply be an unfolding of higher- or lower-order functions in another hierarchy.

A hierarchical structure may not be reversible. We can see this by looking at processes of planning. This type of planning is simply and graphically demonstrated in a relevant way in space travel. Launching a manned craft and returning it to its starting point is a two-point boundary problem. Different considerations are involved in getting it from one point to the second than in getting it from the second to the first. In the forward process, high velocity and the effect of gravity are the critical factors. It is important to know how many *g*'s are exerted on the body. In the backward process, air resistance and the need for parachutes or other deceleration devices, heat transfer, and the tolerance of the heat shield material are the important factors. These factors are also there in the launching process but they are not the critical ones. The two sets of factors must be taken into consideration to solve the entire problem.

Another feature of the two hierarchies of the forward and backward planning processes is that they cannot be easily linked. To see this we note that both processes can be represented by "exploding trees." The theory we briefly describe here enables us to contain or limit the size of the hierarchy to its essential components and still get an effective representation of a system. The result is to facilitate the process of interaction between the two hierarchies.

There is a dichotomy as to whether a hierarchy is a convenient tool of the mind or whether nature actually is endowed with hierarchical structures and functions. Considerable evidence has been put together to support the latter idea. Here are brief eloquent expressions in defense of each point of view.

> The immense scope of hierarchical classification is clear. It is the most powerful method of classification used by the human brain-mind in ordering experience, observations, entities and information. Though not yet definitely established as such by neurophysiology and psychology, hierarchical classification probably represents the prime mode of coordination or organization (i) of cortical processes, (ii) of their mental correlates, and (iii) of the expression of these in symbolisms and languages. The use of hierarchical ordering must be as old as human thought, conscious and unconscious .... (Whyte, 1969).

Direct confrontation of the large and the small is avoided in nature through the use of a hierarchical linkage. Bigness is avoided by functionally bounding the ratio between the size of the hierarchy and that of its levels.

> In the more than fifty years of my intimate preoccupation, with the phenomena and problems of morphogenesis . . . I have been unable to find a way of deriving, free from all preoccupations, a comprehensive and realistic description of the developmental process otherwise than by reference to a dualistic concept, according to which the discrete units are enmeshed in, and interplay with, an organized reference system of unified dynamics of the collective of which they are the members. (Weiss, 1971).

In either case, our present purpose is to assist our understanding of the interrelations that at least the model claims exist.

Now, for some properties of hierarchical structures:

(1) A significant observation is that they usually consist of a few kinds of subsystems in various combinations and arrangements—a multitude of proteins from about 20 amino acids; a very large variety of molecules from about a hundred elements.

(2) They are nearly decomposable; i.e., connections between levels are far simpler than the connections between the elements in a level. Thus, only the aggregate properties of the level determine the interactions between levels and not the properties of the individual elements.

(3) Regularities in the interactions between levels may themselves be classified and coded taking advantage of redundancy in complex hierarchical structures to obtain greater simplicity in explanation. Thus, for example, the trajectory of a system over an entire period of time may be simply described in terms of the differential law generating that trajectory at individual instants of time.

Advantages of hierarchies:

(1) They provide a meaningful integrations of systems. The integrated behavior or function of a hierarchical organization accounts for the fact that complicated changes in a large system can result in a single component. It is the opposite of what we generally expect.

(2) They use aggregates of element in the form of levels to accomplish tasks.

(3) Greater detail occurs down the hierarchy levels; greater depth in understanding its purpose occurs up the hierarchy levels. From the upper level the constraints of the hierarchy are taken for granted and the question is, "How could the constraints arise?"

(4) Hierarchies are efficient and will evolve in natural systems much more rapidly than nonhierarchic systems having the same number of elements. This is demonstrated below.

(5) Hierarchies are reliable and flexible. Local perturbation does not perturb the entire hierarchy. The overall purpose of the hierarchy is divided among the levels whereby each solves a partial problem and the totality meets the overall purpose. The units on the higher level are not concerned with the overall purpose but with specific goals of that system should be attempted not in terms of the overall goal but in terms of specific goals of each level.

5. Formal Hierarchies

The laws characterizing different levels of a hierarchy are generally different. The levels differ in both structure and function. The proper functioning of a higher level depends on the proper functioning of the lower levels. The basic problem with a hierarchy is to seek understanding at the highest levels from interactions of the various levels of the hierarchy rather than directly from elements of the levels. At this state of development of the theory the choice of levels in a hierarchy generally depends on the knowledge and interpretation of the observer. Let us note in passing that, for example, the optics eigenvectors are estimates of the inverse square natural law. The approach will represent laws characterizing problems of greater complexity.

DEFINITION 4. An ordered set is any set S with a binary relation $\leqslant$ which satisfies the reflexive, antisymmetric, and transitive laws:

Reflexive:	For all x, $x \leqslant x$;
Antisymmetric:	If $x \leqslant y$ and $y \leqslant x$, then $x = y$;
Transitive:	If $x \leqslant y$ and $y \leqslant z$, then $x \leqslant z$.

For any relation $x \leqslant y$ (read, y includes x) of this type, we may define $x < y$ to mean that $x \leqslant y$ and $x \neq y$. y is said to cover (dominate) x if $x < y$ and if $x < t < y$ is possible for no t.

Ordered sets with a finite number of elements can be conveniently represented by directed graphs. Each element of the system is represented by a vertex so that an arc is directed from a to b if $b < a$.

DEFINITION 5. A simply or totally ordered set (also called a chain) is an ordered set with the additional property that if $x, y \in S$ then either $x \leqslant y$ or $y \leqslant x$.

DEFINITION 6. A subset E of an ordered set S is said to be bounded from above if there is an element $s \in S$ such that $x \leqslant s$ for every $x \in E$. The element s is called an *upper bound* of E. We say E has a supremum or least upper bound in S if E has upper bounds and if the set of upper bounds U has an element u_1 such that $u_1 \leqslant u$ for all $u \in U$. The element u_1 is unique and is called the supremum of E in S. The symbol sup is used to represent a supremum. (For finite sets the largest elements and the upper bounds are the same.)

Similar definitions may be given for sets bounded from below, a *lower bound* and *infimum*. The symbol inf is used.

There are many ways of defining a hierarchy. The one which suits our needs best here is the following:

> We use the notation $x^- = \{y \mid x \text{ covers } y\}$ and $x^+ = \{y \mid y \text{ covers } x\}$, for any element x in an ordered set.

DEFINITION 7. Let H be a finite partially ordered set with largest element b. H is a *hierarchy* if it satisfies the conditions

(a) There is a partition of H into sets L_k, $k = 1,\dots, h$, where $L_1 = \{b\}$.

(b) $x \in L_k$ implies $x^- \subset L_{k+1}$, $k = 1,\dots, h - 1$.

(c) $x \in L_k$ implies $x^+ \subset L_{k-1}$, $k = 2,\dots, h$.

For each $x \in H$, there is a suitable weighting function (whose nature depends on the phenomenon being hierarchically structured):

$$w_x\colon x^- \to [0, 1] \text{ such that } \sum_{y \in x^-} w_x(y) = 1.$$

The sets L_i are the *levels* of the hierarchy, and the function w_x is the *priority function* of the elements in one level with respect to the objective x. We observe that even if

$x^- \neq L_k$ (for some level L_k), w_x may be defined for all of L_k by setting it equal to zero for all elements in L_k not in x^-.

The weighting function, we feel, is a significant contribution toward the application of hierarchy theory.

DEFINITION 8. A hierarchy is *complete* if, for all $x \in L_k$ $x^+ = L_{k-1}$, for $k = 2,\ldots, h$.

We can state the central question:

BASIC PROBLEM. Given any element $x \in L_\alpha$, and subset $S \subset L_\beta$ $(\alpha < \beta)$, how do we define a function $w_{x,S}$: $S \to [0, 1]$ which reflects the properties of the priority functions w_y on the levels L_k, $k = \alpha,\ldots, \beta - 1$. Specifically, what is the function $w_{b,L_h}: L_h \to [0, 1]$.

In less technical terms, this can be paraphrased thus:

Given a social (or economic) system with a major objective b, and the set L_h of basic activities such that the system can be modeled as a hierarchy with largest element b and lowest level L_h, what are the priorities of the elements of L_h with respect to b?

From the standpoint of optimization, allocating a resource among the elements any interdependence must also be considered. Analytically, interdependence may take the form of input–output relations such as, for example, the interflow of products between industries. A high-priority industry may depend on the flow of matrial from a low-priority industry. In an optimization framework, the priority of the elements enables one to define the objective function to be maximized, and other hierarchies supply information regarding constraints, e.g., input–output relations. An application of this exists.

We now present our method for solving the Basic Problem. Assume that

$$Y = \{y_1,\ldots, y_{m_k}\} \in L_k$$

and that $X = \{x_1,\ldots, x_{m_{k+1}}\} \in L_{k+1}$. (Observe that according to the remark following Definition 7, we may assume that $Y = L_k$, $X = L_{k+1}$. Also assume that there is an element $z \in L_{k-1}$, such that $y \subset z^-$. We then consider the priority functions

$$w_z\colon Y \to [0, 1] \qquad \text{and} \qquad w_y\colon X \to [0, 1], \qquad j = 1,\ldots, n_k.$$

We construct the "priority function of the elements in X with respect to z," denoted w, w: $X \to [0, 1]$, by

$$w(x_i) = \sum_{j=1}^{n_k} w_{y_j}(x_i)\, w_z(y_j), \qquad i = 1,\ldots, n_{k+1}.$$

It is obvious that this is no more than the process of weighting the influence of the element y_j on the priority of x_i by multiplying it with the importance of y_i with respect to z.

The algorithms involved are simplified if one combines the $w_{y_j}(x_i)$ into a matrix B by setting $b_{ij} = w_{y_j}(x_i)$. If, further, we set $W_i = w(x_i)$ and $W_j' = w_z(y_j)$, then the above formula becomes

$$W_i = \sum_{j=1}^{n_k} b_{ij} W_j', \qquad i = 1,\ldots, n_{k+1}.$$

Thus, we may speak of the *priority vector* W and, indeed, of the *priority matrix* B of the $(k + 1)$st level; this gives the final formulation

$$W = BW'.$$

The foregoing may be summarized in a principle.

Principle of hierarchical composition. Given two finite sets S and T, let S be a set of properties and let T be a set of objects which have the properties as characteristics. Assume that a numerical weight, priority, or index of relative importance, $w_j > 0$, $j = 1,..., n$, is associated with each $s_j \in S$, such that $\sum_{j=1}^{n} w_j = 1$. Let $w_{ij} > 0$, $i = 1,..., m$, with $\sum_{i=1}^{m} w_{ij} = 1$, be weights associated with $t_i \in T$, $i = 1,..., m$, relative to s_j. Then the convex combination of w_{ij}, $j = 1,..., n$,

$$\sum_{j=1}^{n} w_{ij} w_j, \qquad i = 1,..., m,$$

gives the numerical priority or relative importance of t_i with respect to S. Note that the principle generalizes to a chain of sets. An axiomatization of the principle of hierarchical composition would be useful.

The following is a first step toward validating the above principle, as it shows that the ordinal preferences are preserved under composition.

DEFINITION 9. Suppose that for each subgoal or activity C_j in L_k there is an ordinal scale o_j over the activities C_α ($\alpha = 1,..., n_{k+1}$) in L_{k+1}. Define a partial order over L_{k+1} by: $C_\alpha \geqslant C_\beta$ if and only if for $j = 1,..., n_k$, $o_{\alpha j} \geqslant o_{\beta j}$.

It is easy to prove:

THEOREM 10. *Let $(w_{1j},..., w_{n_{k+1}j})$ be the eigenvector for L_{k+1} with respect to C_j, and suppose it preserves the order of the $o_{\alpha j}$. Let $W_1,..., W_{n_{k+1}}$ be the (composite) priority vector for L_{k+1}. Then $C_\alpha \geqslant C_\beta$ implies $W_\alpha \geqslant W_\beta$.*

Thus hierarchical composition preserves ordinal preference.

The following is easy to prove:

THEOREM 11. *Let H be a complete hierarchy with largest element b and h levels. Let B_k be the priority matrix of the kth level, $k = 2,..., h$. If W' is the priority vector of the pth level with respect to some element z in the $(p - 1)$st level, then the priority vector W of the qth level $(p < q)$ with respect to z is given by*

$$W = B_q B_{q-1} \cdots B_{p+1} W',$$

Thus, the priority vector of the lowest level with respect to the element b is given by:

$$W = B_h B_{h-1} \cdots B_2 W'.$$

If L_1 has a single element, as usual, W' is just a scalar; if more, a vector.

The following observation holds for a complete hierarchy but it is also useful in general. The priority of an element in a level is the sum of its priorities in each of the comparison subsets to which it belongs; each weighted by the fraction of elements of the level which belong to that subset and by the priority of that subset. The resulting set of priorities of the elements in the level is then normalized by dividing by its sum. The priority of a subset in a level is equal to the priority of the dominating element in the next level.

The reader will have noted that our earlier definition of a hierarchy is more general than we needed in a complete hierarchy. We have made some observations here on general hierarchies but we have used extensively in practice the general definition given. Any of the following applications would require a longer discussion than given here and several of them are in the process of publication.

Examples

1. *School Selection*

Three high schools, A, B, C, were analyzed according to their desirability from the standpoint of a candidate. Six characteristics were selected for the comparison. They are: learning, friends, school life, vocational training, college preparation, and music classes. The pairwise judgment matrices (See Table 10 and the matrices below) were filled by the author's teen-age son and wife in a substantial debate.

TABLE 10

Comparison of Characteristics with Respect to Overall Satisfaction with School

	Learning	Friends	School life	Vocational training	College preparation	Music classes
Learning	1	4	3	1	3	4
Friends	$\frac{1}{4}$	1	7	3	$\frac{1}{5}$	1
School life	$\frac{1}{3}$	$\frac{1}{7}$	1	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{6}$
Vocational training	1	$\frac{1}{3}$	5	1	1	$\frac{1}{3}$
College preparation	$\frac{1}{3}$	5	5	1	1	3
Music classes	$\frac{1}{4}$	1	6	3	$\frac{1}{3}$	1

Comparison of Schools with Respect to the Six Characteristics

Learning

	A	B	C
A	1	$\frac{1}{3}$	$\frac{1}{2}$
B	3	1	3
C	2	$\frac{1}{3}$	1

Friends

	A	B	C
A	1	1	1
B	1	1	1
C	1	1	1

School life

	A	B	C
A	1	5	1
B	$\frac{1}{5}$	1	$\frac{1}{5}$
C	1	5	1

Vocational training

	A	B	C
A	1	9	7
B	$\frac{1}{9}$	1	$\frac{1}{5}$
C	$\frac{1}{7}$	5	1

College preparation

	A	B	C
A	1	$\frac{1}{2}$	1
B	2	1	2
C	1	$\frac{1}{2}$	1

Music classes

	A	B	C
A	1	6	4
B	$\frac{1}{6}$	1	$\frac{1}{3}$
C	$\frac{1}{4}$	3	1

The eigenvector of the matrix in Table 10 given by:

$$(0.32, 0.14, 0.03, 0.13, 0.24, 0.14)$$

and its corresponding eigenvalue is $\lambda_{\max} = 7.49$, which is far from the consistent value 6. No revision of the matrix was made. Normally such inconsistency would indicate that we should reconsider the arrangements.

The eigenvalues and eigenvectors of the other six matrices are given in Table 11.

TABLE 11

Eigenvalues and Eigenvectors of the Comparison Matrices of the Schools with Respect to the Characteristics

$\lambda_{\max} = 3.05$ Learning	$\lambda_{\max} = 3$ Friends	$\lambda_{\max} = 3$ School life	$\lambda_{\max} = 3.21$ Vocational training	$\lambda_{\max} = 3$ College preparation	$\lambda_{\max} = 3.05$ Music classes
0.16	0.33	0.45	0.77	0.25	0.69
0.59	0.33	0.09	0.05	0.50	0.09
0.25	0.33	0.46	0.17	0.25	0.22

To obtain the overall ranking of the schools, we multiply the last matrix on the right by the transpose of the vector of weights of the characteristics. This yields:

$$A = 0.37, \qquad B = 0.38, \qquad C = 0.25.$$

The individual want to school A because it had almost the same rank as school B; yet school B was a private school charging close to $1600 a year and school A was free.

This is an example where we were able to bring in a lower-priority item, e.g., the cost of the school, to add to the argument that A is favored by the candidate. The actual hierarchy is shown in Fig. 4.

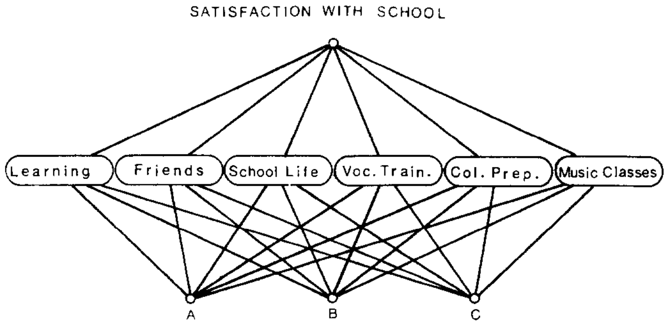

FIG. 4. School satisfaction hierarchy.

2. *Psychotherapy*

The hierarchical method of prioritization may be used to provide insight into psychological problem areas in the following manner: Consider an individual's overall well-being as the single top level entry in a hierarchy. Conceivably this level is primarily affected by childhood, adolescent, and adult experiences. Factors in growth and maturity which impinge upon well-being may be the influences of the father and the mother separately as well as their influences together as parents, the socioeconomic background; sibling relationships, one's peer group, schooling, religious status, and so on.

The factors above, which comprise the second level in our hierarchy, are further affected by criteria pertinent to each. For example, the influence of the father may be broken down to include his temperament, strictness, care, and affection. Sibling relationships can be further characterized by the number, age differential, and sexes of siblings; peer pressure and role modeling provide a still clearer picture of the effects of friends, schooling, and teachers.

As an alternative framework of description for the second level, we might include self-respect, security, adaptability to new people and new circumstances, and so on, influencing or as influenced by the elements above.

A more complete setting for a psychological history might include several hundreds of elements at each level, chosen by trained individuals and placed in such a way as to derive the maximum understanding of the subject in question.

Here we consider a highly restricted form of the above, where the individual in question feels that his self-confidence has been severely undermined and his social adjustments have been impaired by a restrictive situation during childhood. He is questioned about his childhood experiences only and is asked to relate the following elements pairwise on each level:

Level I	Overall well-being (O.W.).
Level II	Self-respect, sense of security, ability to adapt to others (R, S, A).

Level III Visible affection shown for subject (V),
Ideas of strictness, ethics (E),
Actual disciplining of child (D),
Emphasis on personal adjustment with others (O).

Level IV Influence of mother, father, both (M, F, B),

The replies in the matrix form were as follows, supplied by Ms X, a student:

O.W.

	R	S	A
R	1	6	4
S	$\frac{1}{6}$	1	3
A	$\frac{1}{4}$	$\frac{1}{3}$	1

R

	V	E	D	O
V	1	6	6	3
E	$\frac{1}{6}$	1	4	3
D	$\frac{1}{6}$	$\frac{1}{4}$	1	$\frac{1}{2}$
O	$\frac{1}{3}$	$\frac{1}{3}$	2	1

S

	V	E	D	O
V	1	6	6	3
E	$\frac{1}{6}$	1	4	3
D	$\frac{1}{6}$	$\frac{1}{4}$	1	$\frac{1}{2}$
O	$\frac{1}{3}$	$\frac{1}{3}$	2	1

A

	V	E	D	O
V	1	$\frac{1}{5}$	$\frac{1}{3}$	1
E	5	1	4	$\frac{1}{5}$
D	3	$\frac{1}{4}$	1	$\frac{1}{4}$
O	1	5	4	1

V

	M	F	B
M	1	9	4
F	$\frac{1}{9}$	1	8
B	$\frac{1}{4}$	$\frac{1}{8}$	1

E

	M	F	B
M	1	1	1
F	1	1	1
B	1	1	1

D

	M	F	B
M	1	9	6
F	$\frac{1}{9}$	1	$\frac{1}{4}$
B	$\frac{1}{6}$	4	1

O

	M	F	B
M	1	5	5
F	$\frac{1}{5}$	1	$\frac{1}{3}$
B	$\frac{1}{5}$	3	1

The eigenvector of the first matrix, a, is given by:

	O, W
R	0.701
S	0.193
A	0.106

The matrix, b, of eigenvectors of the second row of matrices is given by:

	R	S	A
V	0.604	0.604	0.127
E	0.213	0.213	0.281
D	0.064	0.064	0.120
O	0.119	0.119	0.463

The matrix, c, of eigenvalues of the third row of matrices is given by:

	V	E	D	O
M	0.721	0.333	0.713	0.701
F	0.210	0.333	0.061	0.097
B	0.069	0.333	0.176	0.202

The final composite vector of influence on well-being obtained from the product cba is given by:

$$\text{Mother} = 0.635,$$
$$\text{Father} = 0.209,$$
$$\text{Both} = 0.156.$$

3. *Choosing a Job.*

A student who had just received his Ph. D. was interviewed for three jobs. His criteria selecting the jobs and their pairwise comparison matrix are given in Table 12. The pairwise comparison matrices of the jobs with respect to each criterion are:

Research

	A	B	C
A	1	$\frac{1}{4}$	$\frac{1}{2}$
B	4	1	3
C	2	$\frac{1}{3}$	1

Growth

	A	B	C
A	1	$\frac{1}{4}$	$\frac{1}{5}$
B	4	1	$\frac{1}{2}$
C	5	2	1

Benefits

	A	B	C
A	1	3	$\frac{1}{3}$
B	$\frac{1}{3}$	1	1
C	3	1	1

Colleagues

	A	B	C
A	1	$\frac{1}{3}$	5
B	3	1	7
C	$\frac{1}{5}$	$\frac{1}{7}$	1

Location

	A	B	C
A	1	1	7
B	1	1	7
C	$\frac{1}{7}$	$\frac{1}{7}$	1

Reputation

	A	B	C
A	1	7	9
B	$\frac{1}{7}$	1	5
C	$\frac{1}{9}$	$\frac{1}{5}$	1

TABLE 12

Overall Satisfaction with Job

	Research	Growth	Benefits	Colleagues	Location	Reputation
Research	1	1	1	4	1	$\frac{1}{2}$
Growth	1	1	2	4	1	$\frac{1}{2}$
Benefits	1	$\frac{1}{2}$	1	5	3	$\frac{1}{2}$
Colleagues	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{5}$	1	$\frac{1}{3}$	$\frac{1}{3}$
Location	1	1	$\frac{1}{3}$	3	1	1
Reputation	2	2	2	3	1	1

The eigenvalue of the matrix of Table 12 is $\lambda_{\max} = 6.35$ and the corresponding eigenvector is

The eigenvalues and eigenvectors of the remaining matrices are given in Table 13. The composite vector for the jobs is given by

$$A = 0.40, \qquad B = 0.34, \qquad C = 0.26.$$

The differences were sufficiently large for the cadidate to accept the offer of job A.

TABLE 13

Eigenvalues and Eigenvectors of the Comparison Matrices of the Jobs with Respect to the Criteria

$\lambda_{\max} = 3.02$ Research	$\lambda_{\max} = 3.02$ Growth	$\lambda_{\max} = 3.56$ Benefits	$\lambda_{\max} = 3.05$ Colleagues	$\lambda_{\max} = 3$ Location	$\lambda_{\max} = 3.21$ Reputation
0.14	0.10	0.32	0.28	0.47	0.77
0.63	0.33	0.22	0.65	0.47	0.17
0.24	0.57	0.46	0.07	0.07	0.05

4. *Selecting a Plan for Vacation*

With a view to spending a week for vacation, four places were evaluated in terms of the following criteria:

F_1 : Cost of the trip from Philadelphia
F_2 : Sight-seeing opportunities
F_3 : Entertainment (doing things)
F_4 : Way of travel
F_5 : Eating places

The places considered were:

S: Short trips (i.e., New York, Washington, Atlantic City, New Hope, etc.)
Q: Quebec
D: Denver
C: California.

The comparison matrix of the criteria with respect to overall satisfaction with a vacation plan is given in Table 14.

TABLE 14

Comparison Matrix of Criteria for Vacation[a]

	Cost	Sight-seeing	Entertain-ment	Way of travel	Eating places	Eigen-vector
Cost	1	$\frac{1}{5}$	$\frac{1}{5}$	1	$\frac{1}{3}$	0.09
Sightseeing	5	1	$\frac{1}{5}$	$\frac{1}{5}$	1	0.13
Entertainment	5	5	1	$\frac{1}{5}$	1	0.23
Way of travel	1	5	5	1	5	0.43
Eating places	3	1	1	$\frac{1}{5}$	1	0.13

[a] $\lambda_{\max} = 6.78$.

The comparison matrices of vacation sites with respect to the criteria are:

Cost

	S	Q	D	C	Eigenvector
S	1	3	7	9	0.58
Q	$\frac{1}{3}$	1	6	7	0.30
D	$\frac{1}{7}$	$\frac{1}{6}$	1	3	0.08
C	$\frac{1}{9}$	$\frac{1}{7}$	$\frac{1}{3}$	1	0.04

$\lambda_{\max} = 4.21$

Sightseeing

	S	Q	D	C	Eigenvector
S	1	$\frac{1}{5}$	$\frac{1}{6}$	$\frac{1}{4}$	0.06
Q	5	1	2	4	0.45
D	6	$\frac{1}{2}$	1	6	0.38
C	4	$\frac{1}{4}$	$\frac{1}{6}$	1	0.12

$\lambda_{\max} = 4.34$

Entertainment

	S	Q	D	C	Eigenvector
S	1	7	7	$\frac{1}{2}$	0.36
Q	$\frac{1}{7}$	1	1	$\frac{1}{7}$	0.06
D	$\frac{1}{7}$	1	1	$\frac{1}{7}$	0.06
C	2	7	7	1	0.52

$\lambda_{\max} = 4.06$

Way of travel

	S	Q	D	C	Eigenvector
S	1	4	$\frac{1}{4}$	$\frac{1}{3}$	0.21
Q	$\frac{1}{4}$	1	$\frac{1}{2}$	3	0.19
D	4	2	1	3	0.41
C	3	$\frac{1}{3}$	$\frac{1}{3}$	1	0.18

$\lambda_{\max} = 5.38$

Eating places

	S	Q	D	C	Eigenvector
S	1	1	7	4	0.43
Q	1	1	6	3	0.38
D	$\frac{1}{7}$	$\frac{1}{6}$	1	$\frac{1}{4}$	0.05
C	$\frac{1}{4}$	$\frac{1}{3}$	4	1	0.14

$\lambda_{\max} = 4.08$

The ranking of the four places obtained by the prioritization is:

$$S = 0.29,$$
$$Q = 0.23,$$
$$D = 0.25,$$
$$C = 0.24.$$

This shows that the four are almost equally preferred, with short-trip places having a slight edge over the others.

5. *Conflict and Planning Applications*

In an interesting application to conflict analysis, the theory has been used to construct a hierarchy whose levels represent the actors who influence or control the outcome of the conflict, the objectives of the actors, their policies, their strategies, and the set of plausible outcomes which can result from their actions. The analysis leads to weights or priorities for the outcomes. The method offers ground for approaching the parties on what may work best when all their combined interests are taken into consideration or to show them where to modify their positions to obtain a jointly more desirable outcome.

An outcome in a planning problem is often referred to as a scenario. To ensure taking into account the widest "plausible" set of possible outcomes (or scenarios) it is desirable to adopt for each actor an outcome which he would like to pursue by himself. The set of outcomes is then hierarchically weighted by the weights of the actors composed with the weights of their objectives and finally with those of their strategies (Saaty and Rogers, 1976). A composite outcome, the resultant of all the influences on the set of outcomes, is obtained. This is the likely or composite future. The likely future is characterized in detail by a set of state variables. The values of these variables are calibrated by weighting the corresponding values of the variables for each individual future considered. These are usually determined on a difference scale according to the strength of differences of each variable from its value in the present outcome taken as the zero reference point. The present is assumed to be the best-known outcome with which other outcomes may be compared. The purpose of such an analysis is to examine the attitudes of the actors about the future within a hierarchical framework which they can help define and to offer them an opportunity to bargain and change their position, hopefully to obtain a more favorable outcome (Alexander and Saaty, 1977).

6. Decomposition and Aggregation or Clustering

There are essentially two fundamental ways in which the idea of a hierarchy can be used.

The first is by now clear: it has to do with modeling the real world hierarchically.

The second is probably even more fundamental than the first and points to the real power of hierarchies in nature. It is to break things down into large groupings or clusters and then break each of these into smaller clusters and so on. The object would then be to obtain the priorities of all the elements by means of clustering. This is by far a more efficient process than treating all the elements together. Thus, it is immaterial whether we think of hierarchies as intrinsic in nature, as some have maintained, or whether we simply use them because of our limited capacity to process information. In either case, they are a very efficient way of looking at complex problems.

To decompose a hierarchy into clusters, one must first decide on which elements to group together in each cluster. This is done according to the proximity or similarity of the elements with respect to the function they perform or property they share and regarding which we need to know the priority of these elements. One must then conduct comparisons on the clusters and on the subclusters and then recompose the clusters to obtain a true reflection of the overall priorities. If this process works, the result after the decomposition should be the same as the result if there were no decomposition. Let us illustrate with the example of distances from Philadelphia of the six cities mentioned earlier.

A Distance Hierarchy

We now structure the example of distances between cities into a hierarchy.

If we group the cities into clusters according near-equivalent distances from Philadelphia, we have three classes compared in the following matrix.

	Chicago Montreal	London San Franciso	Cairo Tokyo	Eigenvector
Chicago Montreal	1	$\frac{1}{7}$	$\frac{1}{9}$	0.056
London San Francisco	7	1	$\frac{1}{4}$	0.26
Cairo Tokyo	9	4	1	0.68

If we now compare the cities in each cluster separately according to their relative distances from Philadelphia, we have, on using for the 2 by 2 case the scale $1 + \epsilon$:

	Chicago	Montreal	Eigenvector
Chicago	1	2	0.67
Montreal	$\frac{1}{2}$	1	0.33

	Cairo	Tokyo	Eigenvector
Cairo	1	1/1.5	0.4
Tokyo	1.5	1	0.6

	San Francisco	London	Eigenvector
San Francisco	1	1/1.3	0.43
London	1.3	1	0.57

Now we multiply the first eigenvector by 0.056, the second by 0.26, and the third by 0.68 to obtain the overall relative distance vector:

Cairo	Tokyo	Chicago	San Franciso	London	Montreal
(0.27	0.41	0.037	0.11	0.15	0.019)

which is a good estimate of the actual relative distance vector

(0.278	0.361	0.032	0.132	0.177	0.019)

Let us assume that we have a set of n elements. If we wish to compare the elements in pairs to obtain a ratio scale ranking by solving the eigenvalue problem, $(n^2 - n)/2$ judgments would be necessary. Suppose now that 7 is the maximum number of elements which can be compared with any reasonable (psychological) assurance of consistency. Then n must be first decomposed into equivalence classes of seven clusters or subsets, each of these decomposed in turn to seven new clusters and so on down generating levels of a hierarchy until we obtain a final decomposition each of whose sets has no more than seven of the orginal elements. Let $\{x\}$ denote the smallest integer greater than or equal to x. We have:

THEOREM 12. *The maximum number of comparisons obtained from the decomposition of a set of $n > 1$ elements into a hierarchy of clusters (under the assumption that no more than seven elements are compared simultaneously), is bounded by* $(7/2)(7^{\{\log n/\log 7\}} - 1)$ *and this bound is sharp.*

Proof. We have the following for the *number of comparisons* in each level where we must have in the hth or last level at most seven elements in each cluster.

1. 0
2. $(7^2 - 7)/2$
3. $7 \times (7^2 - 7)/2$

$\vdots$

k. $7^{h-2} \times (7^2 - 7)/2, \qquad \text{where} \quad 7^{h-2} \times 7 = n, \quad h = \{\log n/\log 7\} + 1, \quad h > 2.$

The sum of these comparisons is

$$21 \times (7^{h-1} - 1)/(7 - 1) = (7/2)(7^{\{\log n/\log 7\}} - 1).$$

To show that the bound is sharp it is sufficient to put $n = 7^m$.

Remark. It looks as if the Saint Ives conundrum finds its solution in hierarchies.

The efficiency of a hierarchy may be defined to be the ratio of the number of direct pairwise comparisons required for the entire set of n elements involved in the hierarchy, as compared with the number of pairwise comparisons resulting from clustering as described above.

THEOREM 13. *The efficiency of a hierarchy is of the order of $n/7$.*

Proof. To prove the theorem we must compare

$$(n^2 - n)/2 \qquad \text{with} \qquad (7/2)(7^{\{\log n/\log 7\}} - 1).$$

Let $n = 7^{m+\epsilon}$, $0 \leqslant \epsilon < 1$. Then we clearly have

$$7^{2m+2\epsilon} - 7^{m+\epsilon}/7 \cdot (7^m - 1) \geqslant 7^{m+\epsilon}/7 = n/7.$$

Thus $n/7$ is equal to the efficiency.

One might naturally ask why we do not use 2 in place of 7 for even greater efficiency. We note that in using a hierarchy we seek both consistency and good correspondence to reality. The former is greater the smaller the size of each matrix; the latter is greater the larger the size of the matrix due to the use of redundant information. Thus we have a trade off. Actually, we have seen that using the consistency index C the number 7 is a good practical bound on n, a last outpost, as far as consistency is concerned.

Remark. The exponential efficiency is $\log_7 n$ or, in general, if we replace 7 by s we have $\log_s n$.

Suppose we have a set of 98 elements to which we want to assign priority. We decompose the problem into seven sets each having on the average 14 elements. Now we cannot compare 14 elements so we decompose each of these sets into two sets each having no more than 7 elements. We then compare the elements among themselves.

To look at the efficiency of this process closely we note that if it were possible to compare 98 elements among themselves, we would require $((98)^2 - 98)/2 = 4753$ comparisons. On the other hand, if we divide them into seven clusters of 14 elements each, and then do pairwise comparisons of the seven clusters, we need $(7^2 - 7)/2 = 21$ comparisons. Each cluster can now be divided into two clusters each with seven elements. Comparing two clusters falling under each of the 14-element clusters requires one comparison but there are seven of these, hence, we require seven comparisons on this level; then then we need $14 \times 21 = 294$ comparisons on the lowest level. The total number of comparisons in this hierarchical decomposition is $21 + 7 + 294 = 322$ as compared with 4753 comparisons without clustering. Indeed the theorem is satisfied since $322 \leqslant 4753/7$.

Clustering a complex problem into hierarchical form has two advantages:

(1) Great efficiency in making pairwise comparisons,

(2) Greater consistency under the assumption of a limited capacity of the mind to compare more than 7 ± 2 elements simultaneously.

The efficiency of a hierarchy has been illustrated by Simon (1962) with an example of two men assembing watches, one by constructing modular or component parts from elementary parts and using these to construct higher-order parts and so on, and the other by assembling the entire watch piece by piece from beginning to end. If the first man is interrupted, he only has to start reassembling a small module but if the second man is interrupted, he has to start reassembling the watch from the beginning. If the watch has 1000 components and the components at each level have 10 parts, the first man will, of course, have to make the components and then from these make subassemblies in a total of 111 operations. If p is the probability of an interruption while a part is being added to an incomplete assembly, then the probability that the first man completes a piece without interruption is $(1 - p)^{10}$ and that for the second is $(1 - p)^{1000}$. For the first man, an interruption would cost the time required to assemble 5 parts. The cost to the second man will, on the average, be the time needed to assemble $1/p$ parts which is approximately the expected number of parts without interruption. If $p = .01$ (a chance in a hundred that either man would be interrupted in adding any 1 part), the cost to the first man is 5 and to the second man 100. The first man will assemble 111 components while the second would make just 1. However, the first man will complete an assembly in $(1 - 0.01)^{-10} = 10/9$ attempts whereas the second man will complete an assembly in $(1 - 0.01)^{-1000} = (1/44) \times 10^6$ attempts. Thus the ratio of the efficiency of the first man to that of the second man is given by:

$$\frac{100/0.99^{1000}}{111\{[(1/0.99^{10}) - 1]5 + 10\}} \approx 2000.$$

In man-made systems, the task of managing a complex enterprise is, in general, considerably simplified when it is broken down into subsystems or levels that are individually more tractable, i.e., with a manager having a limited span of management. The steps of solving a large-scale problem are more simply and efficiently accomplished when they are modularized, e.g., by taking n sets of m variables each, than when mn variables are taken simultaneously.

7. Comments on Relation to Other Work

In his summary paper, Shepard (1972) indicates that reasearch on dominance matrices and corresponding measurement has not been as extensive as research on the other three types: proximity, profile, and conjoint. We are essentially interested in dominance matrices and their use in deriving ratio scales and, furthermore, in the measurement of hierarchical impacts. Let us compare the method with work done by others. We hope that we may be forgiven if our comparison is not as complete as some may like to see. As it is, the core of the ideas was improvised and grew completely out of applications. Then it had to be integrated into the main stream of the literature.

Thurstone's model (1927) of comparative judgment demands pairwise comparison of the objects, but only to the extent that one is more preferred to or greater than another. He recovers information over the stimuli by impossing assumptions of normality on the judgmental process. Under additional assumptions on the parameters, e.g., equal variances or zero covariances, he recovers various "metric" information on the stimuli. A number of restrictions are associated with Thurstone's approach. For example, Guilford (1928) recommends limiting the range of probabilities that one stimulus is judged to be more than another.

Torgerson (1958) has systematized and extended Thurstone's method for scaling; in particular, concentrating on the case in which covariance terms are constant, correlation terms equal, and distributions homoscedastic, i.e., they have equal variances.

Luce and Suppes (1964) and Suppes and Zinnes (1963) have proposed what Coombs (1964) calls the Bradley–Terry–Luce (BTL) model using the logistic curve which is a log transform of the probability distribution. Although this is different from assuming normality, in practice it is difficult to distinguish between the BTL model and the case in Thurstone's work where he assumes normal distributions and equal variances. The BTL model is more rigorously grounded in a theory of choice behavior. Coombs discusses the essential distinction between the two models.

We can contrast our assumptions with psychometric tradition. We do not begin with the supposition that ratio judgments are independent probabilistic processes. Instead, we investigate the consequence of changes in the judgments through perturbations on the entire set of judgments. This type of approach leads to the criterion of consistency. Thus, obtaining solutions in our method is not a statistical procedure.

Briefly, many psychometric methods perform aggregation of judgments in the course of solving for a scale. We assume that if there is aggregation of judgments, it occurs prior to the ratio estimate between two stimuli. Therefore, our solution procedure is not concerned with assumptions of distributions of judgments. However, if we want to compare any solution with the criterion of consistency, we appeal to statistical reasoning and perturbations over the entire matrix of judgments.

Our use of metric information in the matrix of subjects' judgments generates strong parallels with principal component analysis, except that the data give dominance rather than similarity or covariance information. In principal component analysis $\lambda_{\max}$ is emphasized, but one also solves for all the λ's. However, the results must be interpreted differently (Hotelling, 1933).

In our analysis the nature of the stimuli and the task presented to subjects are also similar to "psychophysical" scaling, as typified by Stevens and Galanter (1964) and recently used widely in many attempts to construct composite measures of political variables including "national power." Stevens' technique imposes consistency by asking the subjects to compare simultaneously each stimulus with all others, producing only one row of our matrix. This means the hypothesis of unidimensionality cannot be tested directly. If Stevens' method is used, one should take care that the judgments over stimuli are known to be consistent or nearly so. In addition, there is no way of relating one scale to another as we do with the hierarchy.

Krantz (1972) has axiomatized alternative processes relating stimuli to judgments and has derived existence theorems for ratio scales. Comparable axiomatization has not been extended to hierarchies of ratio scales.

Some people have approached problems of scaling as if the cognitive space of stimuli were inherently multidimensional, but we choose instead to decompose this multidimensional structure hierarchically in order to establish a quantitative as well as qualitative relation among dimensions. The individual dimensions in multidimensional scaling solutions functionally resemble individual eigenvectors on any one level of our hierarchy.

The formal problem of constructing a scale as the normalized eigenvector w in the equation $Aw = \lambda w$, for λ a maximum, is similar to extracting the first principal component. When subjects are asked to fill the cells of only one row or one column and the other cells are computed from these (to insure "perfect consistency") the first eigenvalue, n, represents 100% of the variance in the matrix. If, however, "perfect consistency" applies to the data except that a normally distributed random component is added to each cell of the matrix, then one's theory of data would lead to principal factor analysis, and a "single-factor" solution would result. Thus, the imposition of perfect consistency by the experimenter produces an uninteresting result of exact scalability, which was assured by the experimental design of single comparisons. In fact, one can see that if the subjects fill only one row or column of the matrix and if the subjects' task is to generate ratios between pairs of stimuli, then the procedure is formally equivalent to having the subjects locate each stimulus along a continuum with a natural zero at one end: this is the "direct-intensity" technique of psychophysical scaling.

There is no simple relationship of the eigenvalue solution to least-squares solutions, although there have been papers (for example, by Eckart and Young (1936). Keller (1962), and Johnson (1963)) concerned with approximating a matrix of data by a matrix of lower rank, minimizing the sum of the square of the differences. In general, the two solutions are the same when we have consistency. A widely accepted criterion for comparison is not known. Thus, it is not clear which is superior. Iterating the eigenvalue procedure helps us approach consistency, which is our preferred criterion.

Tucker (1958) presents a method for the "determination of parameters of a functional relation by factor analysis." He states, however, that "the rotation of axes problem remains unsolved...," that is, the factor analysis determines the parameters only within a linear transformation. Cliff (1975) suggests methods for the determination of such transformations where a priori theoretical analysis or observable quantities provide a criterion toward which to rotate the arbitrary factor solution.

The hierarchical composition is an inductive generalization of the following idea. We are given weights of elements in one level. We generate a matrix of column eigenvectors of the elements in the level immediately below this level with respect to each element in this level. Then we use the vector of (weights of) elements in this level to weight the corresponding column eigenvectors. Multiplying the matrix of eigenvectors with the column vector of weights gives the composite vector of weights of the lower-level elements.

Because the matrix of eigenvectors is not an orthogonal transformation, in general the result cannot be interpreted as a rotation. In fact, we are multiplying a vector in the unit n-simplex by a stochastic matrix. The result is another vector in the unit simplex. Algebraists have often pointed to a distinction between problems whose algebra has a structural geometric interpretation and those in which algebra serves as a convenient method for doing calculations. Statistical methods have a convenient geometric interpretation. Perturbation methods frequently may not.

In the works of Hammond and Summers (1965) concern is expressed regarding the performance of subjects in situations involving both linear and nonlinear relations among stimuli before concluding that the process of inductive inference is primarily linear. In our model, subjects' responses to linear and nonlinear cues seem to be adequately captured by the pairwise scaling method described here, by using the hierarchical decomposition approach in order to aggregate elements which fall into comparability classes according to the possible range of the scale used for the comparison.

Note that our solution of the information integration problem discussed by Anderson (1974) is approached through an eigenvalue formulation which has a linear structure. However, the scale defined by the eigenvector itself is a highly nonlinear function of the data. The process by means of which the eigenvector is generated involves complex addition, multiplication, and averaging. To perceive this complexity one may examine the eigenvector as a limiting solution of the normalized row sums of powers of the matrix.

Anderson (1974) also makes a strong point that validation of a response scale ought to satisfy a criterion imposed by the algebraic judgment model. Such a criterion in our case turns out to be consistency.

Finally, it may be useful to mention briefly a graph-theoretic approach to consistency. A directed graph on n vertices which is complete (i.e., every pair of vertices is connected by a directed arc) is called a tournament. It can be used to represent dominance pairwise comparisons among n objects. Its cycles would then represent intransitivity. For example, every three vertices define a triangle, but not all triangles form 3-cycles. The number of cycles of given length is used to define an intransitivity index for that order, e.g., between triples or quadruples. Inconsistency is then defined (see Marshall (1971)) in terms of the ratio of the number of three or four or more cycles in a given graph to the maximum number of cycles of that order. For 3-cycles the maximum number is $(n^3 - n)/24$ for n odd, and $(n^3 - 4n)/24$ for n even. For 4-cycles it is $(n^3 - n)(n - 3)/48$ for n odd, and $(n^3 - 4n)(n - 3)/48$ for n even. These results have not been generalized to k-cycles. However, the average number of k-cycles for a random orientation of the arcs of a complete graph is $(k - 1)!\,\binom{n}{k}(\frac{1}{2})^k$. As yet we have found no relationship between this definition of inconsistency and our eigenvalue-related definition. It is not likely that there will be. The above 3-cycle result is due to M. G. Kendall together with its statistical implications. It is nicely discussed in standard statistical references (see, for example, Moroney (1968)).

8. Conclusions

There are a number of possible extensions of the use of the ideas discussed here that are being presently pursued. In a recent Ph. D. dissertation Peter Blair has applied the ideas to the design of an energy park using hierarchies together with input–output analysis. He used goal programming to ensure the real consistency of flows of materials to conform with the priority weights assigned them. In another dissertation R. Mariano used the ideas to structure problems for energy rationing. Three applications have been made recently. The first was to design a national transport plan for the Sudan, and the second used the approach of analyzing attitudes toward the growth of a large Mexican corporation by working with its planners. The third involved work with the U.S. Navy to help improve the management of one of its important units by interacting with the admiral in charge.

Applications to problems of risk are of particular interest to us at present in the analysis of "limits to growth" for the National Science Foundation.

In conclusion, we have found modeling complex problems with hierarchies useful in stimulating participation and interaction among the people concerned. It seems to provide an opportunity for richer involvement of decision makers both in the formulation and in the quantitatively oriented solution of their problems. We feel that there is substantial work yet to be done on the theory, but enough is known to make it useful even now. We recently extended the idea of consistency to hierarchies and their stabilities.

ACKNOWLEDGMENTS

I am grateful to my friends J. P. Bennett and D. H. Krantz for many helpful comments and for their encouragement, and to the referee for excellent suggestions.

REFERENCES

ALEXANDER, J. M., & SAATY, T. L. The forward and backward processes of conflict analysis. *Behavioral Science*, 1977, March.

ANDERSON, N. H. Information integration theory: a brief survey. In D. H. Krantz, R. C. Atkinson, R. D. Luce, and P. Suppes (Eds.), *Contemporary developments in mathematical psychology*, Vol. II. San Francisco: Freeman, 1974.

CLIFF, N. Complete orders from incomplete data: interactive ordering and tailored testing. *Psychological Bulletin*, 1975, **82**, No. 2, 289–302.

COOMBS, C. H. *A theory of data*. New York/London/Sydney: Wiley, 1964.

ECKART, C., & YOUNG, G. The approximation of one matrix by another of lower rank. *Psychometrika*, 1936, **1**, No. 3, 211–218.

GANTMACHER, F. R. *The theory of matrices*, Vol. II. New York: Chelsea, 1960, pp. 53 and 60.

GUILFORD, J. P. The method of paired comparisons as a psychometric method. *Psychological Review*, 1928, **35**, 494–506.

HAMMOND, K. R., & SUMMERS, D. A. Cognitive dependence on linear and nonlinear cues. *Psychological Review*, 1965, **72**, No. 3, 215–224.

HOTELLING, H. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 1933, **24**, 417–441, 498–520.

JOHNSON, R. M. On a theorem stated by Eckart and Young. *Psychometrika*, Sept. 1963, **28**, No. 3, 259–263.

KELLER, J. B. Miscellanea, factorization of matrices by least-squares. *Biometrika*, 1962, **49**, 1 and 2, 239–242.

KRANTZ, D. H. A theory of magnitude estimation and cross modality matchings. *Journal of Mathematical Psychology*, 1972, **9**, 168–199.

KRANTZ, D. H. Fundamental measurement of force and Newton's first and second laws of motion. *Philosophy of Science*, Dec. 1973, **40**, No. 4, 481–495.

LUCE, R. D., & SUPPES, P. Preference, utility and subjective probability. In *Handbook of mathematical psychology*, Vol. III, 1964.

MARSHALL, C. W. *Applied graph theory*. New York: Wiley-Interscience, 1971.

MILLER, G. A. The magical number seven plus or minus two: some limits on our capacity for processing information. *The Psychological Review*, March 1956, **63**, 81–97.

MORONEY, J. J. *Facts from figures*. Baltimore: Penguin, 1968.

Pattee, H. H. (Ed.). *Hierarchy theory, the challenge of complex systems.* New York: George Braziller, 1973.

Saaty, T. L. Hierarchies and priorities—eigenvalue analysis. University of Pennsylvania, 1975.

Saaty, T. L., & Khouja, M. A measure of world influence. *Peace Science,* June 1976.

Saaty, T. L., & Rogers, P. C. The future of higher education in the United States (1985–2000). *Socio-Economic Planning Sciences,* Dec. 1976, **10**, No. 6, 251–264.

Shepard, R. N. A taxonomy of some principal types of data and of multidimensional methods for their analysis. In R. N. Shepard, A. K. Romney, and S. B. Nerlove (Eds.), *Multidimensional scaling: theory and applications in the behavioral sciences,* Vol. I. New York: Seminar Press, 1972, pp. 21–47.

Simon, H. A., & Ando, A. Aggregation of variables in dynamic systems. *Econometrica,* April 1961, **29**, No. 2, 111–138.

Simon, H. A. The architecture of complexity. *Proceedings of the American Philosophical Society,* Dec. 1962, **106**, 467–482.

Stevens, S. Measurement, psychophysics, and utility. In C. W. Churchman and P. Ratoosh (Eds.), *Measurement, definitions and theories.* New York: Wiley, 1959.

Stevens, S., & Galanter, E. Ratio scales and category scales for a dozen perceptual continua. *Journal of Experimental Psychology,* 1964, **54**, 377–411.

Suppes, P., & Zinnes, J. L. Basic measurement theory. In R. D. Luce, R. R. Bush, E. Galanter (Eds.), *Handbook of mathematical psychology,* Vol. I, 1963.

Thurstone, L. L. A law of comparative judgment. *Psychological Review,* 1927, **34**, 273–286.

Torgerson, W. S. *Theory and methods of scaling.* New York: Wiley, 1958.

Tucker, L. R. Determination of parameters of a functional relation by factor analysis. *Psychometrika,* March 1958, **23**, No. 1.

Weiss, P. A. *Hierarchically organized systems in theory and practice.* New York: Hafner, 1971.

Whyte, L. L., Wilson, A. G., & Wilson, D. (Eds.). *Hierarchical structures,* Amsterdam/New York: Elsevier, 1969.

Wielandt, H. Unzerlegbare, nicht negative matrizen. *Mathematische Zeitschrift,* 1950, **52**, 642–648.

Wilkinson, J. J. *The algebraic eigenvalue problem,* Chap. 2. London: Oxford Univ. Press (Clarendon), 1965.

Received: May 17, 1976