

## Chapter 2

# The Seven Pillars of the Analytic Hierarchy Process

### 2.1 Introduction

The Analytic Hierarchy Process (AHP) provides the objective mathematics to process the inescapably subjective and personal preferences of an individual or a group in making a decision. With the AHP and its generalization, the Analytic Network Process (ANP), one constructs hierarchies or feedback networks, then makes judgments or performs measurements on pairs of elements with respect to a controlling element to derive ratio scales that are then synthesized throughout the structure to select the best alternative.

Fundamentally, the AHP works by developing priorities for alternatives and the criteria used to judge the alternatives. Usually the criteria, whose choice is at the mercy of the understanding of the decision-maker (irrelevant criteria are those that are not included in the hierarchy), are measured on different scales, such as weight and length, or are even intangible for which no scales yet exist. Measurements on different scales, of course, cannot be directly combined. First, priorities are derived for the criteria in terms of their importance to achieve the goal, then priorities are derived for the performance of the alternatives on each criterion. These priorities are derived based on pairwise assessments using judgment, or ratios of measurements from a scale if one exists. The process of prioritization solves the problem of having to deal with different types of scales, by interpreting their significance to the values of the user or users. Finally, a weighting and adding process is used to obtain overall priorities for the alternatives as to how they contribute to the goal. This weighting and adding parallels what one would have done arithmetically prior to the AHP to combine alternatives measured under several criteria having the *same* scale (a scale that is often common to several criteria is money) to obtain an overall result. With the AHP a multidimensional scaling problem is thus transformed to a unidimensional scaling problem.

The seven pillars of the AHP are: (1) **Ratio scales, proportionality, and normalized ratio scales** are central to the generation and synthesis of priorities, whether in the AHP or in any multicriteria method that needs to integrate existing ratio scale measurements with its own derived scales; in addition, ratio scales are the only way to generalize a decision theory to the case of dependence and feedback because ratio scales can be both multiplied, and added—when they belong to the same scale such as a priority scale; when two judges arrive at two different ratio scales for the same problem one needs to test the compatibility of their answers and accept or reject their closeness. The AHP has a non-statistical index for doing this. Ratio scales can also be used to make decisions within an even more general framework involving several hierarchies for benefits, costs, opportunities and risks, and using a common criterion such as *economic* to ensure commensurability; ratio scales are essential in proportionate resource allocation as in linear programming, recently generalized to deal with relative measurement for both the objective function and the constraints obtaining a ratio scale solution vector form which it is possible to decide on the relative values of the allocated resources; one can associate with each alternative a vector of benefits, costs, time of completion, etc., to determine the best alternative subject to all these general concerns; (2) **Reciprocal paired comparisons** are used to express judgments semantically automatically linking them to a numerical fundamental scale of absolute numbers (derived from stimulus response relations) from which the principal eigenvector of priorities is then derived; the eigenvector shows the dominance of each element with respect to the other elements; an element that does not have a particular property is automatically assigned the value zero in the eigenvector without including it in the comparisons; dominance along all possible paths is obtained by raising the matrix to powers and normalizing the sum of the rows; inconsistency in judgment is allowed and a measure for it is provided which can direct the decision maker in both improving judgment and arriving at a better understanding of the problem; scientific procedures for giving less than the full set of  $n(n - 1)/2$  judgments in a matrix have been developed; using interval judgments eventually leading to the use of optimization and statistical procedures is a complex process which is often replaced by comparing ranges of values of the criteria, performing sensitivity analysis, and relying on conditions for the insensitivity of the eigenvector to perturbations in the judgments; the judgments may be considered as random variables with probability distributions; the AHP has at least three modes for arriving at a ranking of the alternatives: (a) **Relative**, which *ranks* a few alternatives by comparing them in pairs and is particularly useful in new and exploratory decisions, (b) **Absolute**, which *rates* an unlimited number of alternatives one at a time on intensity scales constructed separately for each covering criterion and is particularly useful in decisions where there is considerable knowledge to judge the relative importance of the intensities and develop priorities for them; if desired, a few of the top rated alternatives can then be compared against each other using the relative mode to obtain further refinement of the priorities; (c) **Benchmarking**, which ranks alternatives by including a known alternative in the group and comparing the other against it; (3) **Sensitivity of the principal right eigenvector** to perturbation in judgments limits the number of elements in each set of

comparisons to a few and requires that they be homogeneous; the left eigenvector is only meaningful as reciprocal; due to the choice of a unit as one of the two elements in each paired comparison to determine the relative dominance of the second element, it is not possible to derive the principal left eigenvector directly from paired comparisons as the dominant element cannot be decomposed a priori; as a result, to ask for how much less one element is than another we must take the reciprocal of what we get by asking how much more the larger element is; (4) **Homogeneity and clustering** are used to extend the fundamental scale gradually from cluster to adjacent cluster, eventually enlarging the scale from 1–9 to 1– $\infty$ ; (5) **Synthesis that can be extended to dependence and feedback** is applied to the derived ratio scales to create a uni-dimensional ratio scale for representing the overall outcome. Synthesis of the scales derived in the decision structure can only be made to yield correct outcomes on known scales by additive weighting. It should be carefully noted that additive weighting in a hierarchical structure leads to a multilinear form and hence is non-linear. It is known that under very general conditions such multilinear forms are dense in general function spaces (discrete or continuous), and thus linear combinations of them can be used to approximate arbitrarily close to any nonlinear element in that space. Multiplicative weighting, by raising the priorities of the alternatives to the power of the priorities of the criteria (which it determines through additive weighting!) then multiplying the results, has four major flaws: (a) It does not give back weights of existing same ratio scale measurements on several criteria as it should; (b) It assumes that the matrix of judgments is always consistent, thus sacrificing the idea of inconsistency and how to deal with it, and not allowing redundancy of judgments to improve validity about the real world; (c) Most critically, it does not generalize to the case of interdependence and feedback, as the AHP generalizes to the Analytic Network Process (ANP), so essential for the many decision problems in which the criteria and alternatives depend on each other; (d) It always preserves rank which leads to unreasonable outcomes and contradicts the many counterexamples that show rank reversals should be allowed; (6) **Rank preservation and reversal** can be shown to occur without adding or deleting criteria, such as by simply introducing enough copies of an alternative or for numerous other reasons; this leaves no doubt that rank reversal is as intrinsic to decision making as rank preservation also is; it follows that any decision theory must have at least two modes of synthesis; in the AHP they are called the distributive and ideal modes, with guidelines for which mode to use; rank can always be preserved by using the ideal mode in both absolute measurement and relative measurement; (7) **Group judgments** must be integrated one at a time carefully and mathematically, taking into consideration when desired the experience, knowledge, and power of each person involved in the decision, without the need to force consensus, or to use majority or other ordinal ways of voting; the theorem regarding the *impossibility* of constructing a social utility function from individual utilities that satisfies four reasonable conditions which found their validity with *ordinal* preferences is *no longer true* when *cardinal* ratio scale preferences are used as in the AHP. Instead, one has the *possibility* of constructing such a function. To deal with a large group requires the use of questionnaires and statistical procedures for large samples.

## 2.2 Ratio Scales

A *ratio* is the relative value or quotient  $a/b$  of two quantities  $a$  and  $b$  of the same kind; it is called commensurate if it is a rational number, otherwise it is incommensurate. A statement of the equality of two ratios  $a/b$  and  $c/d$  is called *proportionality*. A ratio scale is a set of numbers that is invariant under a similarity transformation (multiplication by a positive constant). The constant cancels when the ratio of any two numbers is formed. Either pounds or kilograms can be used to measure weight, but the ratio of the weight of two objects is the same for both scales. An extension of this idea is that the weights of an entire set of objects whether in pounds or in kilograms can be standardized to read the same by normalizing. In general if the readings from a ratio scale are  $aw_i^*$ ,  $i = 1, \dots, n$ , the standard form is given by  $w_i = aw_i^*/aw_i^* = w_i^*/w_i^*$  as a result of which we have  $w_i = 1$ , and the  $w_i$ ,  $i = 1, \dots, n$ , are said to be normalized. We no longer need to specify whether weight for example is given in pounds or in kilograms or in another kind of unit. The weights (2.21, 4.42) in pounds and (1, 2) in kilograms, are both given by (1/3, 2/3) in the standard ratio scale form.

The relative ratio scale derived from a pairwise comparison reciprocal matrix of judgments is derived by solving:

$$\sum_{j=1}^n a_{ij} w_j = \lambda_{\max} w_i \quad (2.1)$$

$$\sum_{i=1}^n w_i = 1 \quad (2.2)$$

with  $a_{ji} = 1/a_{ij}$  or  $a_{ij} a_{ji} = 1$  (the reciprocal property),  $a_{ij} > 0$  (thus  $A$  is known as a positive matrix) whose solution, known as the principal right eigenvector, is normalized as in (2.2). A relative ratio scale does not need a unit of measurement.

When  $a_{ij} a_{jk} = a_{ik}$ , the matrix  $A = (a_{ij})$  is said to be consistent and its principal eigenvalue is equal to  $n$ . Otherwise, it is simply reciprocal. The general eigenvalue formulation given in (2.1) is obtained by perturbation of the following consistent formulation:

$$Aw = \begin{matrix} & A_1 & \cdots & A_n \\ A_1 & \begin{bmatrix} w_1 \\ w_1 \\ \vdots \\ w_n \end{bmatrix} & & \\ \vdots & & \ddots & \\ A_n & \begin{bmatrix} w_1 \\ w_n \\ \vdots \\ w_n \end{bmatrix} & & \end{matrix} = n \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix} = nw.$$

where  $A$  has been multiplied on the right by the transpose of the vector of weights  $w = (w_1, \dots, w_n)$ . The result of this multiplication is  $nw$ . Thus, to recover the scale from the matrix of ratios, one must solve the problem  $Aw = nw$  or  $(A - nI)w = 0$ . This is a system of homogeneous linear equations. It has a nontrivial solution if and

only if the determinant of  $A - nI$  vanishes, that is,  $n$  is an eigenvalue of  $A$ . Now  $A$  has unit rank since every row is a constant multiple of the first row. Thus all its eigenvalues except one are zero. The sum of the eigenvalues of a matrix is equal to its trace, that is, the sum of its diagonal elements. In this case the trace of  $A$  is equal to  $n$ . Thus  $n$  is an eigenvalue of  $A$ , and one has a nontrivial solution. The solution consists of positive entries and is unique to within a multiplicative constant.

The discrete formulation given in (2.1) and (2.2) above generalizes to the continuous case through Fredholm's integral equation of the second kind and is given by:

$$\int_a^b K(s, t)w(t)dt = \lambda_{\max} w(s) \quad (2.3)$$

$$\lambda \int_a^b K(s, t)w(t)dt = w(s) \quad (2.4)$$

$$\int_a^b w(s)ds = 1 \quad (2.5)$$

where instead of the matrix  $A$  we have as a positive kernel,  $K(s, t) > 0$ . Note that the entries in a matrix depend on the two variables  $i$  and  $j$  which assume discrete values. Thus the matrix itself depends on these discrete variables, and its generalization, the kernel function also depends on two (continuous) variables. The reason for calling it kernel is the role it plays in the integral, where without knowing it we cannot determine the exact form of the solution. The standard way in which (2.3) is written is to move the eigenvalue to the left hand side which gives it the reciprocal form. In general, by abuse of notation, one continues to use the symbol  $\lambda$  to represent the reciprocal value. Our equation for response to a stimulus is now written in the standard form (2.4) with the normalization condition (2.5). Here also, we have the reciprocal property (2.6) and as in the finite case, the kernel  $K(s, t)$  is consistent if it satisfies the relation (2.7):

$$K(s, t)K(t, s) = 1 \quad (2.6)$$

$$K(s, t)K(t, u) = K(s, u), \text{ for all } s, t, \text{ and } u \quad (2.7)$$

An example of this type of kernel is  $K(s, t) = e^{s-t} = e^s/e^t$ . It follows by putting  $s = t = u$ , that  $K(s, s) = 1$  for all  $s$  which is analogous to having ones down the diagonal of the matrix in the discrete case. A value of  $\lambda$  for which Fredholm = sequection has a nonzero solution  $w(t)$  is called a characteristic value (or its reciprocal is called an eigenvalue) and the corresponding solution is called an eigenfunction. An eigenfunction is determined to within a multiplicative constant. If  $w(t)$  is an eigenfunction corresponding to the charateristic value  $\lambda$  and if  $C$  is an arbitrary constant, we can easily see by substituting in the equation that  $Cw(t)$  is also an eigenfunction corresponding to the same  $\lambda$ . The value  $\lambda = 0$  is not

a characteristic value because we have the corresponding solution  $w(t) = 0$  for every value of  $t$ , which is the trivial case, excluded in our discussion.

It may be useful to recount a little of the history of how Fredholm's equation came about in the ratio scale formulation of the AHP. My student Hasan Ait-Kaci and I first recognized the connection between Fredholm's equation and the AHP in a paper we wrote in the late 1970s. In the early 1980s, I and my friend and colleague, Professor Luis Vargas, used this formulation in the framework of neural firing and published several papers on the subject. In December of 1996, I had the nagging idea that the ratio scale relation for electrical firing was not reflected in our solution, and that periodicity had to be involved in the solution with which I began. Many researchers on the brain had considered neural firing in the framework of a damped periodic oscillator. It was my friend Janos Aczel, the leading functional equation mathematician in the world, who provided me with a variety of solutions for the functional equation ( $w(as) = bw(s)$ ). I had proved in the theorem given below that this equation characterizes the solution of Fredholm's equation and its solution is an eigenfunction of that equation. My work is an extension of the work I had done earlier with Vargas. The solution has the form of a damped periodic oscillator of period one. It has an additional logarithmic property that corresponds to Fechner's law discussed later in this paper.

A matrix is consistent if and only if it has the form  $A = (w_i/w_j)$  which is equivalent to multiplying a column vector that is the transpose of  $(w_1, \dots, w_n)$  by the row vector  $(1/w_1, \dots, 1/w_n)$ . As we see below, the kernel  $K(s, t)$  is separable and can be written as

$$K(s, t) = k_1(s)k_2(t).$$

**Theorem**  $K(s, t)$  is consistent if and only if it is separable of the form:

$$K(s, t) = k(s)/k(t). \quad (2.8)$$

**Theorem** If  $K(s, t)$  is consistent, the solution of (2.4) is given by

$$w(s) = \frac{k(s)}{\int_s k(s) ds}. \quad (2.9)$$

In the discrete case, the normalized eigenvector was independent of whether all the elements of the pairwise comparison matrix  $A$  are multiplied by the same constant  $a$  or not, and thus we can replace  $A$  by  $aA$  and obtain the same eigenvector. Generalizing this result we have:

$$K(as, at) = aK(s, t) = k(as)/k(at) = a k(s)/k(t)$$

which means that  $K$  is a homogeneous function of order one. In general, when  $f(ax_1, \dots, ax_n) = a^n f(x_1, \dots, x_n)$  holds,  $f$  is said to be homogeneous of order  $n$ . Because  $K$  is a degenerate kernel, we can replace  $k(s)$  above by  $k(as)$  and obtain  $w(as)$ . We

have now derived from considerations of ratio scales the following condition to be satisfied by a ratio scale:

**Theorem** *A necessary and sufficient condition for  $w(s)$  to be an eigenfunction solution of Fredholm's equation of the second kind, with a consistent kernel that is homogeneous of order one, is that it satisfy the functional equation*

$$w(as) = bw(s)$$

where  $b = \alpha a$ .

We have for the general damped periodic response function  $w(s)$ ,

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

where  $P$  is periodic of period 1 and  $P(0) = 1$ .

We can write this solution as

$$v(u) = C_1 e^{-bu} P(u)$$

where  $P(u)$  is periodic of period 1,  $u = \log s / \log a$  and  $\log ab / -\beta$ ,  $\beta > 0$ . It is interesting to observe the logarithmic function appear as part of the solution. It gives greater confirmation to the Weber–Fechner law developed in the next section.

## 2.3 Paired Comparisons and the Fundamental Scale

Instead of assigning two numbers  $w_i$  and  $w_j$  and forming the ratio  $w_i/w_j$  we assign a single number drawn from the fundamental 1–9 scale of absolute numbers to represent the ratio  $(w_i/w_j)/1$ . It is a nearest integer approximation to the ratio  $w_i/w_j$ . The derived scale will reveal what the  $w_i$  and  $w_j$  are. This is a central fact about the relative measurement approach of the AHP and the need for a fundamental scale.

In 1846 Weber found, for example, that people while holding in their hand different weights, could distinguish between a weight of 20 g and a weight of 21 g, but could not if the second weight is only 20.5 g. On the other hand, while they could not distinguish between 40 and 41 g, they could between 40 and 42 g, and so on at higher levels. We need to increase a stimulus  $s$  by a minimum amount  $\Delta s$  to reach a point where our senses can first discriminate between  $s$  and  $s + \Delta s$ .  $\Delta s$  is called the just noticeable difference (jnd). The ratio  $r = \Delta s/s$  does not depend on  $s$ . Weber's law states that change in sensation is noticed when the stimulus is increased by a constant percentage of the stimulus itself. This law holds in ranges where  $\Delta s$  is small when compared with  $s$ , and hence in practice it fails to hold when  $s$  is either too small or too large. Aggregating or decomposing stimuli as needed into clusters or hierarchy levels is an effective way for extending the uses of this law.

In 1860 Fechner considered a sequence of just noticeable increasing stimuli. He denotes the first one by  $s_0$ . The next just noticeable stimulus is given by

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

based on Weber's law.

Similarly  $s_2 = s_1 + \Delta s_1 = s_1(1 + r) = s_0(1 + r)^2 \equiv s_0 \alpha^2$ . In general  $s_n = s_{n-1} \alpha = s_0 \alpha^n (n = 0, 1, 2, \dots)$ .

Thus stimuli of noticeable differences follow sequentially in a geometric progression. Fechner noted that the corresponding sensations should follow each other in an arithmetic sequence at the discrete points at which just noticeable differences occur. But the latter are obtained when we solve for  $n$ . We have  $n = \frac{(\log s_n - \log s_0)}{\log \alpha}$  and sensation is a linear function of the logarithm of the stimulus. Thus if  $M$  denotes the sensation and  $s$  the stimulus, the psychophysical law of Weber–Fechner is given by

$$M = a \log s + b, \quad a \neq 0.$$

We assume that the stimuli arise in making pairwise comparisons of relatively comparable activities. We are interested in responses whose numerical values are in the form of ratios. Thus  $b = 0$ , from which we must have  $\log s_0 = 0$  or  $s_0 = 1$ , which is possible by calibrating a unit stimulus. Here the unit stimulus is  $s_0$ . The next noticeable stimulus is  $s_1 = s_0 \alpha = \alpha$  which yields the second noticeable response  $a \log \alpha$ . The third noticeable stimulus is  $s_2 = s_0 \alpha^2$  which yields a response of  $2a \log \alpha$ . Thus we have for the different responses:

$$M_0 = a \log s_0, \quad M_1 = a \log \alpha, \quad M_2 = 2a \log \alpha, \dots, \quad M_n = na \log \alpha.$$

While the noticeable ratio stimulus increases geometrically, the response to that stimulus increases arithmetically. Note that  $M_0 = 0$  and there is no response. By dividing each  $M_i$  by  $M_1$  we obtain the sequence of absolute numbers 1, 2, 3, ... of the fundamental 1–9 scale. Paired comparisons are made by identifying the less dominant of two elements and using it as the unit of measurement. One then determines, using the scale 1–9 or its verbal equivalent, how many times more the dominant member of the pair is than this unit. In making paired comparisons, we use the nearest integer approximation from the scale, relying on the insensitivity of the eigenvector to small perturbations (discussed below). The reciprocal value is then automatically used for the comparison of the less dominant element with the more dominant one. Despite the foregoing derivation of the scale in the form of integers, someone might think that other scale values would be better, for example using 1.3 in the place of 2. Imagine comparing the magnitude of two people with respect to the magnitude of one person and using 1.3 for how many there are instead of 2.

We note that there may be elements that are closer than 2 on the 1–9 scale, and we need a variant of the foregoing. Among the elements that are close, we select the smallest. Observe the incremental increases between that smallest one and the rest of the elements in the close group. We now consider these increments to be new elements and pairwise compare them on the scale 1–9. If two of the increments are themselves closer than 2 we treat them as identical, assigning a 1 (we could carry this on ad infinitum—but we will not). In the end each component of the eigenvector of



comparisons of the increments is added to unity to yield the un-normalized priorities of the close elements for that criterion. Note that only the least of these close elements is used in comparisons with the other elements that can be compared directly using the normal 1–9 scale. Its priority is used to multiply the priorities of these close elements and finally the priorities of all the elements are re-normalized.

How large should the upper value of the scale be? Qualitatively, people have a capacity to divide their response to stimuli into three categories: high, medium and low. They also have the capacity to refine this division by further subdividing each of these intensities of responses into high, medium and low, thus yielding in all nine subdivisions. It turns out, from the requirement of homogeneity developed below, that to maintain stability, our minds work with a few elements at a time. Using a large number of elements in one matrix leads to greater inconsistency.

## 2.4 Sensitivity of the Principal Eigenvector Places a Limit on the Number of Elements and Their Homogeneity

To a first order approximation, perturbation  $\Delta w_1$  in the principal eigenvector  $w_1$  due to a perturbation  $\Delta A$  in the matrix  $A$  where  $A$  is consistent is given by:

$$\Delta w_1 = \sum_{j=2}^n (v_j^T \Delta A w_1 / (\lambda_1 - \lambda_j) v_j^T w_j) w_j.$$

The eigenvector  $w_1$  is insensitive to perturbation in  $A$ , if the principal eigenvalue  $\lambda_1$  is separated from the other eigenvalues  $\lambda_j$ , here assumed to be distinct, and none of the products  $v_j^T w_j$  of left and right eigenvectors is small. We should recall that the nonprincipal eigenvectors need not be positive in all components, and they may be complex. One can show that all the  $v_j^T w_j$  are of the same order, and that  $v_1^T w_1$ , the product of the normalized left and right principal eigenvectors is equal to  $n$ . If  $n$  is relatively small and the elements being compared are homogeneous, none of the components of  $w_1$  is arbitrarily small and correspondingly, none of the components of  $v_1^T$  is arbitrarily small. Their product cannot be arbitrarily small, and thus  $w$  is insensitive to small perturbations of the consistent matrix  $A$ . The conclusion is that  $n$  must be small, and one must compare homogeneous elements. Later we discuss placing a limit on the value of  $n$ .

## 2.5 Clustering and Using Pivots to Extend the Scale from 1–9 to 1– $\infty$

In Fig. 2.1, an unripe cherry tomato is eventually and indirectly compared with a large watermelon by first comparing it with a small tomato and a lime, the lime is then used again in a second cluster with a grapefruit and a honey dew where we then divide by the weight of the lime and then multiply by its weight in the first cluster, and then use the honey dew again in a third cluster and so on. In the end we have a

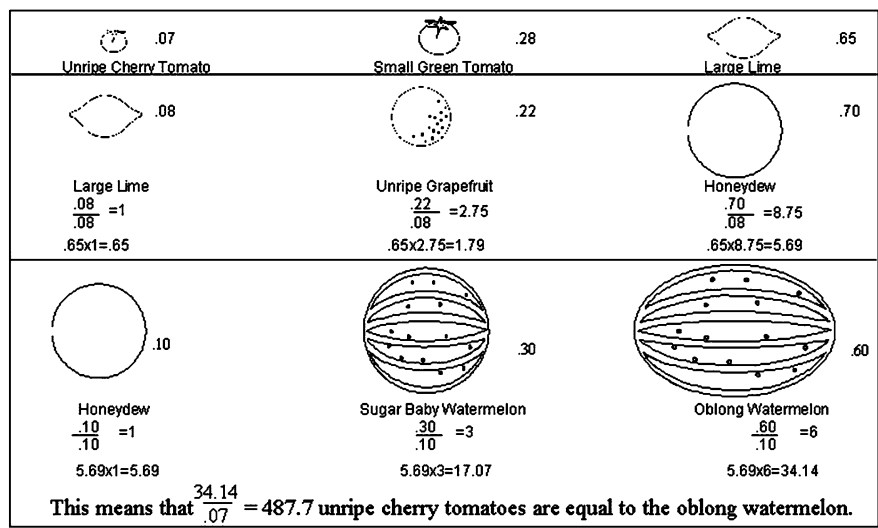
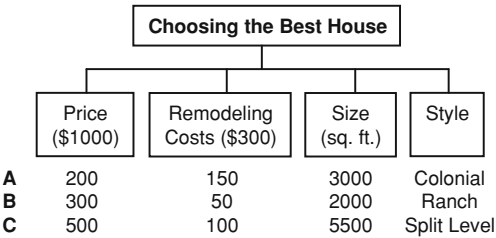


Fig. 2.1 Clustering

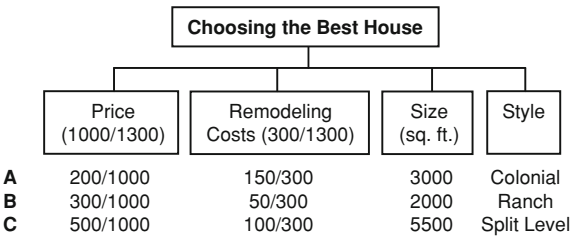
comparison of the unripe cherry tomato with the large watermelon and would accordingly extended the scale from 1–9 to 1–721.

Such clustering is essential, and must be done separately for each criterion. We should note that in most decision problems, there may be one or two levels of clusters and conceivably it may go up to three or four adjacent ranges of homogeneous elements (Maslow put them in seven groupings). Very roughly we have in decreasing order of importance: (1) Survival, health, family, friends and basic religious beliefs some people were known to die for; (2) Career, education, productivity and lifestyle; (3) Political and social beliefs and contributions; (4) Beliefs, ideas, and things that are flexible and it does not matter exactly how one advocates or uses them. Nevertheless one needs them, such as learning to eat with a fork or a chopstick or with the fingers as many people do interchangeably. These categories can be generalized to a group, a corporation, or a government. For very important decisions, two categories may need to be considered. Note that the priorities in two adjacent categories would be sufficiently different, one being an order of magnitude smaller than the other, that in the synthesis, the priorities of the elements in the smaller set have little effect on the decision. We do not have space to show how some *undesirable* elements can be compared among themselves and gradually extended to compare them with *desirable* ones as above. Thus one can go from negatives to positives but keep the measurement of the two types positive, by eventually clustering them separately.

**Fig. 2.2** Ranking houses on four criteria



**Fig. 2.3** Normalization of the measurements



2.6 Synthesis: How to Combine Tangibles With Intangibles—Additive Versus Multiplicative

Let  $H$  be a complete hierarchy with  $h$  levels. Let  $B_k$  be the priority matrix of the  $k$ th level,  $k = 2, \dots, h$ . If  $W'$  is the global priority vector of the  $p$ th level with respect to some element  $z$  in the  $(p - 1)$ st level, then the priority vector  $W$  of the  $q$ th level ( $p < q$ ) with respect to  $z$  is given by the multilinear (and thus non-linear) form,

$$W = B_q B_{q-1} \dots B_{p+1} W'.$$

The global priority vector of the lowest level with respect to the goal is given by,

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

In general,  $W' = 1$ . The sensitivity of the bottom level alternatives with respect to changes in the weights of elements in any level can be studied by means of this multilinear form.

Assume that a family is considering buying a house and there are three houses to consider A, B, and C. Four factors dominate their thinking: the price of the house, the remodeling costs, the size of the house as reflected by its footage and the style of the house which is an intangible. They have looked at three houses with numerical data shown below on the quantifiables (Fig. 2.2):

If we add the costs on price and modeling and normalize we obtain respectively (A, B, C) = (0.269, 0.269, 0.462). Now let us see what is needed for normalization to yield the same result.

First we normalize for each of the quantifiable factors. Then we must normalize the factors measured with respect to a single scale (Fig. 2.3).

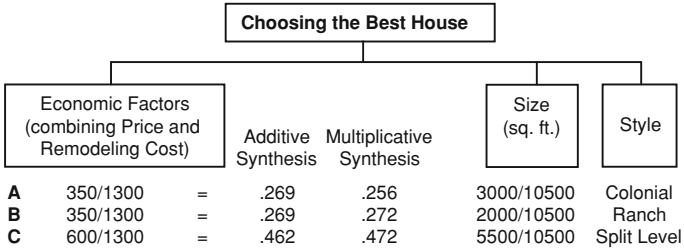


Fig. 2.4 Combining the two costs through additive or multiplicative syntheses

Here we learn two important lessons to be used in the general approach. Normalizing the alternatives for the two criteria involving money in terms of the money involved on both criteria leads to relative weights of importance for the criteria. Here for example Price is in the ratio of about three to one when compared with Remodeling Cost and when compared with the latter with respect to the goal of choosing the best house, it is likely to be assigned the value “moderate” which is nearly three times more as indicated by the measurements. Here the criteria Price and Remodeling Cost derive their priorities only from the alternatives because they are equally important factors, although they can also acquire priorities from higher level criteria as to their functional importance with respect to the ease and availability of different amounts of money. We now combine the two factors with a common scale by weighting and adding. We have (Fig. 2.4):

The left column and its decimal values in the second column give the exact value of the normalized dollars spent on each house obtained by additive synthesis (weighting and adding). By aggregating the two factor measured with dollars into a single factor, one then makes the decision as to which house to buy by comparing the three criteria as to their importance with respect to the goal.

The second lesson is that *when the criteria have different measurements, their importance cannot be determined from the bottom up through measurement of the alternatives, but from the top down, in terms of the goal. The same process of comparison of the criteria with respect to the goal is applied to all criteria if, despite the presence of a physical scale, they are assumed to be measurable on different scales as they might when actual values are unavailable or when it is thought that such measurement does not reflect the relative importance of the alternatives with respect to the given criterion. Imagine that no physical scale of any kind is known!* We might note in passing that the outcome of this process of comparison with respect to higher level criteria yields meaningful (not arbitrary) results as noted by two distinguished proponents of multi-attribute value theory (MAVT) Buede and Maxwell [1], who wrote about their own experiments in decision making:

These experiments demonstrated that the MAVT and AHP techniques, when provided with the same decision outcome data, very often identify the same alternatives as ‘best’. The other techniques are noticeably less consistent with MAVT, the Fuzzy algorithm being the least consistent.

Multiplicative synthesis, as in the third column of numbers above, done by raising each number in the two columns in the previous table to the power of its criterion measured in the relative total dollars under it, multiplying the two outcomes for each alternative and normalizing, *does not* yield the *exact answer* obtained by adding dollars! In addition, A and B should have the same value, but they do not with multiplicative synthesis. The multiplicative “solution” devised for the fallacy of always preserving rank and avoiding inconsistency fails, because it violates the most basic of several requirements mentioned in the introduction to this paper.

Multiplicative and additive syntheses are related analytically through approximation. If we denote by  $a_i$  the priority of the  $i$ th criterion,  $i = 1, \dots, n$ , and by  $x_i$ , the priority of alternative  $x$  with respect to the  $i$ th criterion, then

$$\begin{aligned} \prod x_i^{a_i} &= \exp \log \prod x_i^{a_i} = \exp \left( \sum \log x_i^{a_i} \right) = \exp \left( \sum a_i \log x_i \right) \\ &\approx 1 + \sum a_i \log x_i \approx 1 + \sum (a_i x_i - a_i) = \sum a_i x_i \end{aligned}$$

If desired, one can include a remainder term to estimate the error. With regard to additive and multiplicative syntheses being close, one may think that in the end it does not matter which one is used, but it does. Saaty and Hu [7] have shown that despite such closeness on every matrix of consistent judgments in a decision, the synthesized outcomes by the two methods not only lead to different final priorities (which can cause a faulty allocation of resources) but more significantly to *different rankings* of the alternatives. For all these problems, but more significantly because it does not generalize to dependence and feedback even with consistency guaranteed, and because of the additive nature of matrix multiplication needed to compute feedback in network circuits to extend the AHP to the ANP, I do not recommend ever using multiplicative synthesis. It can lead to an undesirable ranking of the alternatives of a decision.

## 2.7 Rank Preservation and Reversal

Given the assumption that the alternatives of a decision are completely independent of one another, can and should the introduction (deletion) of new (old) alternatives change the rank of some alternatives without introducing new (deleting old) criteria, so that a less preferred alternative becomes most preferred? Incidentally, how one prioritizes the criteria and subcriteria is even more important than how one does the alternatives which are themselves composites of criteria. Can rank reverse among the criteria themselves if new criteria are introduced? Why should that not be as critical a concern? The answer is simple. In its original form utility theory assumed that criteria could not be weighted and the only important elements in a decision were the alternatives and their utilities under the various criteria. Today utility theorists imitate the AHP by rating, and some even by comparing the criteria, somehow. There was no concern then about what would happen to the ranks of the alternatives should the criteria weights themselves

change as there were none. The tendency, even today, is to be unconcerned about the theory of rank preservation and reversal among the criteria themselves.

The house example of the previous section teaches us an important lesson. If we add a fourth house to the collection, the priority weights of the criteria Price and Remodeling Cost would change accordingly. Thus the measurements of the alternatives and their number, which we call structural factors, always affect the importance of the criteria. When the criteria are incommensurate and their functional priorities are determined in terms of yet higher level criteria or goals, one must still weight such functional importance of the criteria by the structural effect of the alternatives. What is significant in all this is that the importance of the criteria always depends on the measurements of the alternatives. If we assume that the alternatives are measured on a different scale for each criterion, it becomes obvious that normalization is the instrument that provides the structural effect to update the importance of the criteria in terms of what alternatives there are. Finally, the priorities of the alternatives are weighted by the priorities of the criteria that depend on the measurements of the alternatives. This implies that the overall ranking of any alternative depends on the measurement and number of all the alternatives. To always preserve rank means that the priorities of the criteria should not depend on the measurements of the alternatives but should only derive from their own functional importance with respect to higher goals. This implies that the alternatives should not depend on the measurements of other alternatives. Thus one way to always preserve rank is to rate the alternatives one at a time. In the AHP this is done through absolute measurement with respect to a complete set of intensity ranges with the largest value intensity value equal to one. It is also possible to preserve rank in relative measurement by using an ideal alternative with full value of one for each criterion.

The logic about what can or should happen to rank when the alternatives *depend* on each other has always been that *anything* can happen. Thus, when the criteria functionally depend on the alternatives, which implies that the alternatives, which of course depend on the criteria, would then depend on the alternatives themselves, rank may be allowed to reverse. The Analytic Network Process (ANP) is the generalization of the AHP to deal with ranking alternatives when there is functional dependence and feedback of any kind. Even here, one can have a decision problem with dependence among the criteria, but with no dependence of criteria on alternatives and rank may still need to be preserved. The ANP takes care of functional dependence, but if the criteria do not depend on the alternatives, the latter are kept out of the supermatrix and ranked precisely as they are dealt with in a hierarchy [8].

Examples of rank reversal abound in practice, and they do not occur because new criteria are introduced. The requirement that rank always be preserved or that it should be preserved with respect to irrelevant alternatives. To every rule or generalization that one may wish to set down about rank, it is possible to find a counterexample that violates that rule. Here is the last and most extreme form of four variants of an attempt to qualify what should happen to rank given by Luce and Raiffa, each of which is followed by a counterexample. They state it but and

then reject it. *The addition of new acts to a decision problem under uncertainty never changes old, originally non-optimal acts into optimal ones. The all-or-none feature of the last form may seem a bit too stringent... a severe criticism is that it yields unreasonable results.* The AHP has a theory and implementation procedures and guidelines for when to preserve rank and when to allow it to reverse. One mode of the AHP allows an irrelevant alternative to cause reversal among the ranks of the original alternatives.

### 2.7.1 Guidelines for Selecting the Distributive or Ideal Mode

The distributive mode of the AHP produces preference scores by normalizing the performance scores; it takes the performance score received by each alternative and divides it by the sum of performance scores of all alternatives under that criterion. This means that with the Distributive mode the preference for any given alternative would go up if we reduce the performance score of another alternative or remove some alternatives.

The Ideal mode compares each performance score to a fixed benchmark such as the performance of the best alternative under that criterion. This means that with the Ideal mode the preference for any given alternative is independent of the performance of other alternatives, except for the alternative selected as a benchmark. Saaty and Vargas [11] have shown by using simulation, that there are only minor differences produced by the two synthesis modes. This means that the decision should select one or the other if the results diverge beyond a given set of acceptable data.

The following guidelines were developed by Millet and Saaty [3] to reflect the core differences in translating performance measures to preference measures of alternatives. *The Distributive (dominance) synthesis mode should be used when the decision maker is concerned with the extent to which each alternative dominates all other alternatives under the criterion. The Ideal (performance) synthesis mode should be used when the decision maker is concerned with how well each alternative performs relative to a fixed benchmark.* In order for dominance to be an issue the decision-maker should regard inferior alternatives as relevant even after the ranking process is completed. This suggests a simple test for the use of the Distributive mode: *if the decision maker indicates that the preference for a top ranked alternative under a given criterion would improve if the performance of any lower ranked alternative was adjusted downward, then one should use the Distributive synthesis mode.* To make this test more actionable we can ask the decision maker to imagine the amount of money he or she would be willing to pay for the top ranked alternative. If the decision maker would be willing to pay more for a top ranked alternative after learning that the performance of one of the lower-ranked alternatives was adjusted downward, then the Distributive mode should be used.

Consider selecting a car: Two different decision makers may approach the same problem from two different points of views even if the criteria and standards are the same. The one who is interested in “getting a well performing car” should use the Ideal

mode. The one who is interested in “getting a car that stands out” among the alternatives purchased by co-workers or neighbors, should use the Distributive mode.

## 2.8 Group Decision Making

Here we consider two issues in group decision making. The first is how to aggregate individual judgments, and the second is how to construct a group choice from individual choices.

### 2.8.1 How to Aggregate Individual Judgments

Let the function  $f(x_1, x_2, \dots, x_n)$  for synthesizing the judgments given by  $n$  judges, satisfy the

1. *Separability condition* (S):  $f(x_1, x_2, \dots, x_n) = g(x_1)g(x_2) \dots g(x_n)$  for all  $x_1, x_2, \dots, x_n$  in an interval  $P$  of positive numbers, where  $g$  is a function mapping  $P$  onto a proper interval  $J$  and is a continuous, associative and cancellative operation. [(S) means that the influences of the individual judgments can be separated as above.]
2. *Unanimity condition* (U):  $f(x, x, \dots, x) = x$  for all  $x$  in  $P$ . [(U) means that if all individuals give the same judgment  $x$ , that judgment should also be the synthesized judgment.]
3. *Homogeneity condition* (H):  $f(ux_1, ux_2, \dots, ux_n) = uf(x_1, x_2, \dots, x_n)$  where  $u > 0$  and  $x_k, ux_k$  ( $k = 1, 2, \dots, n$ ) are all in  $P$ . [For ratio judgments (H) means that if all individuals judge a ratio  $u$  times as large as another ratio, then the synthesized judgment should also be  $u$  times as large.]
4. *Power conditions* ( $P_p$ ):  $f(x_1^p, x_2^p, \dots, x_n^p) = f^p(x_1, x_2, \dots, x_n)$ . [( $P_2$ ) for example means that if the  $k$ th individual judges the length of a side of a square to be  $x_k$ , the synthesized judgment on the area of that square will be given by the square of the synthesized judgment on the length of its side.]

Special case ( $R = P_{-1}$ ):  $f(1/x_1, 1/x_2, \dots, 1/x_n) = 1/f(x_1, x_2, \dots, x_n)$ . [(R) is of particular importance in ratio judgments. It means that the synthesized value of the reciprocal of the individual judgments should be the reciprocal of the synthesized value of the original judgments.]

Aczel and Saaty (see [9, 10]) proved the following theorem:

**Theorem** *The general separable (S) synthesizing functions satisfying the unanimity (U) and homogeneity (H) conditions are the geometric mean and the root-mean-power. If moreover the reciprocal property (R) is assumed even for a single  $n$ -tuple  $(x_1, x_2, \dots, x_n)$  of the judgments of  $n$  individuals, where not all  $x_k$  are equal, then only the geometric mean satisfies all the above conditions.*



In any rational consensus, those who know more should, accordingly, influence the consensus more strongly than those who are less knowledgeable. Some people are clearly wiser and more sensible in such matters than others, others may be more powerful and their opinions should be given appropriately greater weight. For such unequal importance of voters not all  $g$ 's in (S) are the same function. In place of (S), the weighted separability property (WS) is now:  $f(x_1, x_2, \dots, x_n) = g_1(x_1)g_2(x_2) \dots g_n(x_n)$ . [(WS) implies that not all judging individuals have the same weight when the judgments are synthesized and the different influences are reflected in the different functions ( $g_1, g_2, \dots, g_n$ ).]

In this situation, Aczel and Alsina (see [9]) proved the following theorem:

**Theorem** *The general weighted-separable (WS) synthesizing functions with the unanimity (U) and homogeneity (H) properties are the weighted geometric mean  $f(x_1, x_2, \dots, x_n) = x_1^{q_1} x_2^{q_2} \dots x_n^{q_n}$  and the weighted root-mean-powers  $f(x_1, x_2, \dots, x_n) = \sqrt[q_1 x_1^\gamma + q_2 x_2^\gamma + \dots + q_n x_n^\gamma]{\gamma}$ , where  $q_1 + q_2 + \dots + q_n = 1$ ,  $q_k > 0$  ( $k = 1, 2, \dots, n$ ),  $\gamma > 0$ , but otherwise  $q_1, q_2, \dots, q_n, \gamma$  are arbitrary constants.*

If  $f$  also has the reciprocal property (R) and for a single set of entries ( $x_1, x_2, \dots, x_n$ ) of judgments of  $n$  individuals, where not all  $x_k$  are equal, then *only the weighted geometric mean* applies. We give the following theorem which is an explicit statement of the synthesis problem that follows from the previous results, and applies to the second and third cases of the deterministic approach:

**Theorem** *If  $x_1^{(i)}, \dots, x_n^{(i)}$   $i = 1, \dots, m$  are rankings of  $n$  alternatives by  $m$  independent judges and if  $a_i$  is the importance of judge  $i$  developed from a hierarchy for evaluating the judges, and hence  $\sum_{i=1}^m a_i = 1$ , then  $\left(\prod_{i=1}^m x_1^{a_i}\right), \dots, \left(\prod_{i=1}^m x_n^{a_i}\right)$  are the combined ranks of the alternatives for the  $m$  judges.*

The power or priority of judge  $i$  is simply a replication of the judgment of that judge (as if there are as many other judges as indicated by his/her power  $a_i$ ), which implies multiplying his/her ratio by itself  $a_i$  times, and the result follows.

The first requires knowledge of the functions which the particular alternative performs and how well it compares with a standard or benchmark. The second requires comparison with the other alternatives to determine its importance.

### 2.8.2 On the Construction of Group Choice from Individual Choices

Given a group of individuals, a set of alternatives (with cardinality greater than (2), and individual ordinal preferences for the alternatives, Arrow proved with his Impossibility Theorem that it is impossible to derive a rational group choice (construct a social choice function that aggregates individual preferences) from ordinal preferences of the individuals that satisfy the following four conditions, i.e., at least one of them is violated:

*Decisiveness:* the aggregation procedure must generally produce a group order.

*Unanimity:* if all individuals prefer alternative A to alternative B, then the aggregation procedure must produce a group order indicating that the group prefers A to B.

*Independence of irrelevant alternatives:* given two sets of alternatives which both include A and B, if all individuals prefer A to B in both sets, then the aggregation procedure must produce a group order indicating that the group, given any of the two sets of alternatives, prefers A to B.

*No dictator:* no single individual preferences determine the group order.

Using the ratio scale approach of the AHP, it can be shown that because now the individual preferences are cardinal rather than ordinal, it is possible to derive a rational group choice satisfying the above four conditions. It is possible because: (a) Individual priority scales can always be derived from a set of pairwise cardinal preference judgments as long as they form at least a minimal spanning tree in the completely connected graph of the elements being compared; and (b) The cardinal preference judgments associated with group choice belong to a ratio scale that represents the relative intensity of the group preferences.

## Bibliography

1. Buede D, Maxwell DT (1995) Rank disagreement: a comparison of multi-criteria methodologies. *J Multi-Criteria Decis Anal* 4:1–21
2. Luce RD, Raiffa H (1957) *Games and decisions*. Wiley, New York
3. Millet I, Saaty TL (1999) On the relativity of relative measures—accommodating both rank preservation and rank reversal in the AHP. *Eur J Oper Res*
4. Peniwati K (1996) The analytic hierarchy process: the possibility for group decision making. In: *Proceedings of the 4th international symposium on the analytic hierarchy process*, Vancouver, Canada. (Obtainable from RWS Publications, 4922 Ellsworth Avenue, Pittsburgh, PA 15213.), pp 202–214
5. Saaty TL (1999–2000) (ed) *Decision making for leaders*. RWS Publications, 4922 Ellsworth Avenue, Pittsburgh, 15213
6. Saaty TL (2000) *The brain, unraveling the mystery of how it works: the neural network process*, RWS Publications, 4922 Ellsworth Avenue, Pittsburgh, PA 15213
7. Saaty TL, Hu G (1998) Ranking by eigenvector versus other methods in the analytic hierarchy process. *Appl Math Lett* 11(4):121–125
8. Saaty TL (1996) *Decision making with dependence and feedback: the analytic network process*. RWS Publications, 4922 Ellsworth Avenue, Pittsburgh, PA 15213
9. Saaty TL (1994) *Fundamentals of decision making and priority theory*. RWS Publications, 4922 Ellsworth Avenue, Pittsburgh, PA 15213
10. Saaty TL (1990) *Multicriteria decision making: the analytic hierarchy process*. RWS Publications, 4922 Ellsworth Avenue, Pittsburgh, PA
11. Saaty TL, Vargas LG (1993) Experiments on rank preservation and reversal in relative measurement. *Math Comput Model* 17(4/5):13–18
12. Vargas LG (1994) Reply to Schenkerman's avoiding rank reversal in AHP decision support models. *Eur J Oper Res* 74:420–425

Models, Methods, Concepts & Applications of the  
Analytic Hierarchy Process

Saaty, Th.L.; Vargas, L.G.

2012, XIV, 346 p., Hardcover

ISBN: 978-1-4614-3596-9