

RATIO SCALES ARE FUNDAMENTAL IN DECISION MAKING

Thomas L. Saaty
University of Pittsburgh

Abstract: Multicriteria decision making depends on the use of numbers and scales to make trade-offs. We examine in detail the types of numerical scales of measurement there are and which ones seem to work better for measurement in a hierarchic model of a complex problem. It is argued that ratio scales play a unique role in that process. This is particularly true when the decision making theory uses multi-level hierarchic structures and feedback between criteria and alternatives as in the Analytic Hierarchy Process (AHP) and the Analytic Network Process (ANP).

Introduction

The ideas discussed in this paper are very important to the understanding and application of measurement methods in decision making. Understanding them should facilitate comprehension of why one has to work hard to develop a viable, general theory to help people make decisions. Multicriteria methods of measurement require the use of numerical scales in some fashion to make it possible to trade off the ranking of alternatives with respect to one criterion against their ranking with respect to another to obtain an overall ranking. Multicriteria methods must also use numerical scales to establish the weight of each criterion itself. It is unlikely that numerical representation can be avoided despite the raging debate about which numerical scales are more faithful and true to people's feelings and conceptions.

One school of thought argues that preference rankings should be derived according to the weakest possible procedures since people are often uncertain and uninformed about their judgments. It is an analyst's obligation, therefore, to make it easy for them to use whatever information they have to make a decision. Even ordinal ranks and optimality criteria can be used in this process. Another school argues that so long as human beings are able to express the intensity of their preferences, they would not be satisfied with a weaker approach because when they examine a decision later with greater experience and concern, they will no longer see it as the best choice. Stronger scales require greater involvement of one's perceptions, sometimes aided by group participation and debate.

An interesting practical question arises: Is the decision chosen by the many methods unique? Can the use and manipulation of different scales lead to alternative, sometimes less desirable, decisions? Such diversity in arriving at the decision reached could make a methodological approach less of a science, and would call for a more comprehensive decision science to help one select the best method to use.

Scales in Multicriteria Decisions

Some procedures attempt to avoid total reliance on numerical representation by examining the dominance of alternatives through mathematical criteria such as Pareto optimality or deductive reasoning in search of a plausible set of best outcomes. The real problem, however, is one of tradeoffs among the criteria yielding outcomes for alternatives that may not be part of the Pareto set. In essence it should be possible to convert deductive reasoning to a scale representation if one wishes to retain all the information regarding dominance. In any case such methods cannot avoid using numbers to weight the criteria and do not skirt the scaling problem.

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

Ratio scales can be derived from numerical judgments as in the AHP. Intrinsically, there are different kinds of ratio scales. There are absolute ratio scales derived from judgments based on an absolute fundamental scale. The accuracy of the ratios depends on the accuracy of the absolute judgments. There are ratio ratio scales derived from ratios of actual measurements. They give back these measurements. There are ordinal ratio scales derived from interval and ordinal judgments. Such scales indicate dominance but without cardinal meaning attached to their ratios. Finally, there are chaotic ratio scales derived from arbitrary numerical judgments. These ratio scales are as meaningless as are the original numbers assigned.

In multicriteria decisions based on interval or ratio scales, it is not reasonable to expect to choose a different scale for measuring the alternatives with respect to each criterion and yet to arrive at a unique overall decision. A way is required to represent the rankings with respect to each criterion on the same underlying ratio scale (priority theory, based on the dominance of preferences) (Saaty, T.L., 1990) or interval scale (utility theory) (Luce, R.D., and Raiffa, H. 1957). The latter is only valid for weighting the alternatives with criteria weights represented by a ratio scale. One cannot measure criteria or goals on an interval scale and then use them for weighting alternatives since one then obtains a product of two interval scales. In any event, there is the problem of intangible criteria and how to construct their measurement. The next fundamental question is: How are numbers introduced into the analysis of decision making?

Numbers can be used to represent the magnitude (cardinal numbers) and order (ordinal numbers) of preferences in two ways. Ordinal numbers merely indicate what comes before and what comes after according to the magnitude of the number, while with cardinal numbers, one can use differences or multiples or ratios to determine the relative magnitudes among the elements being ranked. The first is by assigning them directly and the second is by deriving a scale which first elicits judgments, perhaps qualitatively, and then represents them or their magnitude numerically by using a fundamental scale. The latter can be represented by a triple: pairs of objects, numbers, and a mapping from pairs of objects to numbers.

The question that is often raised about the direct use of numbers to represent judgments and generate a scale is: How are these numbers conceived in the mind of an individual and associated with judgments and where do they come from? A technical objection to the direct assignment of numbers is that simply claiming they are cardinals does not make it true. Furthermore, even when they can be shown to be

cardinal numbers, one makes all sorts of errors in providing the right ones which rigorous decision theory tries to avoid. Direct assignment is tantamount to guessing, which is not a systematic procedure, where we learn by making an assignment and then revising the judgments.

In the second method, there is usually a way to allow interaction between the derived scale and the fundamental scale, to successively improve both the assignment of numbers and the scale derived from them. Here consistency, which is a necessary but not a sufficient condition for capturing a valid scale, can be used to detect inaccuracies in the numbers assigned and improve the values assigned step by step, if knowledge is available. To ensure sufficiency requires a valid representation of the real world by structuring the decision problem in such a way that the assignment of numbers from a fundamental scale to preferences falls within the capability of the individual to assess relations with a minimum of large errors. This can be accomplished by some techniques, such as the Analytic Hierarchy Process (AHP), by using hierarchic structures with clusters of homogeneous elements. That allows for expressing preference for an alternative over another alternative that is a perturbation of it. Homogeneity ensures that the perturbation is not too large. Thus, deriving scales provides an excellent opportunity to think more deeply about the problem as one improves both consistency and validity. Referring to the argument made earlier, the fundamental scale could and should consist of absolute numbers from which weaker ratio or interval scales can be derived. Here again ordinals are not suitable for a fundamental scale because they do not bring cardinal information that is essential for making tradeoffs.

Relative and Absolute Measurement in the AHP

There is yet one more observation that must be made about methodology. In multicriteria decision theory, scales must be derived; they are not generally already out there to be used. Even when we are dealing with concepts that have physical measurement, we must interpret these measurements in some justifiable mathematical way to make a decision. Measurement on derived scales falls in two general types. We note here what psychologists have observed, namely that people have the ability to apply two ways to compare alternatives. The first is by matching them against norms established through experience and learning, and stored in memory. A new object is judged to measure up or not measure up to these norms by a little or by much. The other way is to compare objects in pairs with respect to a common property to determine which has more of that property. These two innate abilities of all people to express judgment have their numerical counterparts in absolute and relative measurement. Relative measurement proceeds by comparing elements according to dominance with respect to a common property in order to derive weights or priorities for them. One makes comparisons of homogeneous alternatives by answering a question such as: Which of two given people is more gentle and how much more? Pairwise comparisons are measured by taking the lesser alternative as the unit with respect to the property and the greater one as a multiple of that unit expressed by an absolute number on the fundamental scale. Even though the number is absolute the measurement is relative for the two people. A ratio scale is then derived from these paired comparisons.

In absolute measurement one assigns weights or magnitudes to the alternatives one at a time according to a scale of intensities such as: excellent, very good, good, poor. The importance or utility of these shades of difference need prior expert knowledge for each criterion to establish and numerically rate the norms and to use them to assign magnitudes to alternatives. These are then weighted by the priorities of the criteria for an overall ranking.

But what do we do when we have no expert knowledge or norms to develop scales for intensities? It is like asking a lay person to grade college student examinations or to rate restaurants. He may not be sufficiently discriminating to rank them one at a time but he can compare them with each other and arrive at a ranking. In planning the future, we do not know enough about the superiority of different alternative outcomes we have not previously experienced to assign them a numerical value on each criterion. Yet we may be able to rank them through comparison for each criterion. Making a decision is more often a learning process in which even the judgment and norms of experts come into question. Here we are compelled to use relative measurement.

The Fundamental Nature of The Concern About Which Scale To Use

If we assume that it is possible to develop a unified theory for multicriteria decisions to include many of the issues encountered in decision making, which scale would one have to use, or should different scales be used?

A unifying theory must go beyond the simplest decision setting in which the importance of the criteria are assumed to be independent from the alternatives and the preferences of the alternatives independent among themselves (preferences of the alternatives always depend on the criteria used to rank them.) Such a decision has the linear hierarchic form shown in Figure 1.

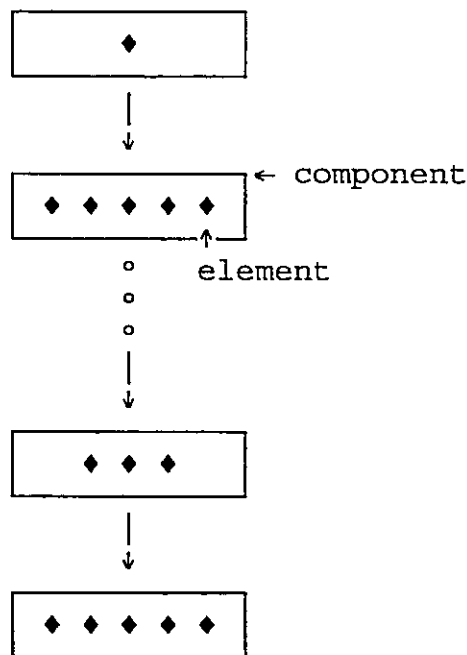


Figure 1: A Linear Hierarchy

In this simple case of a hierarchy, there is wide agreement that ratio scales are needed to develop weights for all the elements involved above the level of alternatives such as time periods, goals, criteria, and subcriteria. In a sense this ranking is even more important than that of the alternatives themselves because the resulting priorities are used to weight the rankings of the alternatives. It appears that the few basic theories (AHP, Utility Theory, Linear and Goal Programming, Bayes Probabilities, and the Electre approach of B. Roy) in MCDM (Multi-Criteria Decision Making) have devoted most of their attention to the method and scale (if they have considered scales at all) used to rank the alternatives without much attention to which scale to use to rank the criteria.

Multicriteria methods for ranking alternatives were established before the development of a rigorous approach for weighting the criteria by ratio scales. Thus, people, for historical and professional reasons, of which they have a commitment to uphold, use one method or another. In light of this new information, it would seem that criteria should always be weighted by ratio scales according to priorities. In the AHP, the same concept of priority is applied to rate the alternatives. In the most noted of these, utility theory, the concept of utility, borrowed from economics, is used to rate the alternatives on an interval scale. The importance of a criterion is not measured directly, but is assumed to register in the weights of the alternatives themselves depending on how widely spread the alternatives are under that criterion. More widely spread alternatives get more weight (using some kind of a scale, such as 1 to 100, from which to draw the multiplying factor). The weights of the alternatives under the various criteria are synthesized

under one of several schemes (which can lead to different rankings because the scheme itself is chosen according to the user's inclination about the importance of the criteria.)

Most decisions, however, involve dependence between and within the categories involved: goals, criteria, and alternatives. For example, in a problem that has no precedent, the criteria can only be observed in the particular alternatives being considered and their derived importance by examining their relative presence in these alternatives. Hence, in this case, the criteria depend on the alternatives. For example, early in its history the effects of AIDS (Acquired Immune Deficiency Syndrome) on an individual could not be described in abstract medical diagnostic terms until they could be observed in the particular patients who were thought to have AIDS. It can also happen that some criteria depend on the alternatives and some do not. Alternatives can depend on other alternatives as when industries perform input-output analysis. A network representation of a decision problem with dependence can have the form shown in Figure 2. In Figure 2 an arrow from a cluster to another indicates the dependence of the elements in the second cluster on those in the first. An arrow from a cluster to itself indicates dependence of the elements among themselves. This dependence takes place with respect to elements (properties or criteria) in the clusters from which arrows emanate towards the given cluster.

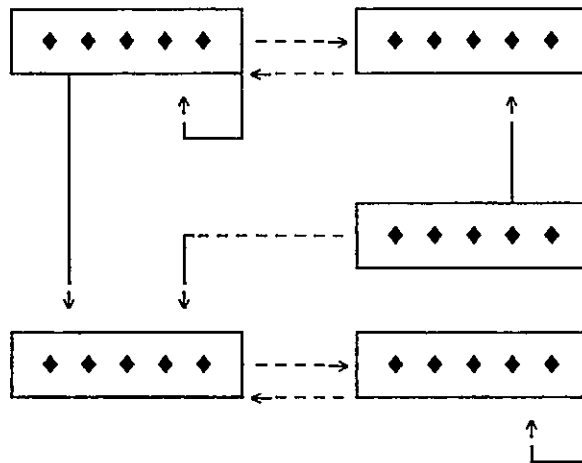


Figure 2: A Non-Linear Network

When the criteria depend on the alternatives, their weights are developed by comparing them with respect to each alternative and these weights are then combined for all the alternatives by using the weights of the alternatives. This weighting and adding cannot be done if the weights of the alternatives are measured on an interval scale, for then the resulting weights of the criteria, would, at a minimum, be measured on an interval scale or something less strong than a ratio scale. This contradicts the requirement for deriving ratio scales to rank the criteria whose priorities are then used to weight the rankings of the alternatives. With this type of dependence consideration no such problem arises if the ranks of the alternatives are measured on a ratio scale.

The problem becomes even more complex when the criteria depend on the alternatives, which Utility Theory does not allow. Should one then use interval scales to rank the alternatives when the criteria are independent from the alternatives and ratio scales when they depend on them? Is not independence a special case of dependence? It is clear that an iterative weighting process would be involved, requiring a scale that allows infinite successive arithmetic operations, which, for that purpose, can only be done on derived ratio scales.

Reasons For Using Ratio Scales

We have seen that ratio scales are always needed in hierarchic structures to rank the criteria and also to rank the alternatives at least in the case of dependence of criteria on alternatives. There are other theoretical and practical reasons (and new observations) that arise in using ratio scales for relative measurement. It is also important to refer to the numerous paradoxes encountered by the utility interval scale approach that weakens its applicability.

Since single ratio scales are essential in hierarchies and the more general case of dependence in decision making, why would one need to introduce interval scales, contrary to the law of parsimony? This principle says that if simple assumptions are sufficient to produce an explanation or deal with a problem, there is no need to seek another explanation through a more complicated route of assumptions. Paradoxes of utility theory (Allais, M 1953; Kahneman, D. and Tversky, A 1979; Kahneman, D., Slovic, P., and Tversky, A. 1982; Luce, R.D. and Raiffa, H. 1957; MacCrimmon, K.R. 1968; Schoemaker, P.J.H. 1980; Tversky, A. 1969) do not occur if ratio scales are used as in the Analytic Hierarchy Process because no assumptions on transitivity and independence from irrelevant alternatives are necessary (Saaty, T.L. 1990).

We noted above that alternatives can be ranked on a ratio scale either through absolute measurement, if there is adequate information to create intensity scales for the criteria, or through relative measurement if there is not. The use of ratio scales has made those using the scale aware of a distinction that needs to be made between absolute and relative measurement, a second reason in favor of ratio scales. Absolute measurement is normative (prescriptive) because of its reliance on standards or norms to rank alternatives but relative measurement is descriptive because it does not require norms. There are times when we may wish to maintain useful standards we have labored hard to establish, such as admitting students to a university and evaluating employee performance under strong conditions of outside competition. At other times, when we have no standards, we still need to make decisions and can if we use relative measurement.

With the foregoing in mind it is important to discuss a third reason for using ratio scales. In using multicriteria decision methods one is concerned with the question: What happens to the rank of alternatives when new ones are added or old ones deleted? It is precisely this issue that separates the use of ratio and of interval scales because it has given rise to a number of paradoxes in the latter.

It had been long held that in a linear decision problem with a ranking of alternatives by absolute measurement, only when criteria are changed in kind or number, or their priorities change, should the rank of alternatives reflect that change. But what happens if we add to the alternatives, copies, near copies, or modifications of an alternative? If we were to always use absolute measurement which ranks the alternatives one at a time, whether on a ratio or an interval scale, in principle, the rank of alternatives would not change when alternatives are added or deleted. This gives rise to a paradox in many decision problems. Too many copies of the same alternative, like having too much of anything, decrease the desirability of that alternative - a situation of abundance. Its opposite, scarcity, can also affect preference. If rank must always be preserved, no matter how many apples are added, an apple continues to be the preferred fruit. This contradicts experience. In this particular case of scarcity and abundance absolute measurement cannot be the only natural way to rank alternatives. It also shows that the number of alternatives and their relative measurements, overlooked in absolute measurement, need to be considered in the operations, not the criteria, of a decision theory because as criteria they strictly imply dependence among the alternatives which is unacceptable. Incidentally, scarcity or uniqueness or whatever would represent manyness cannot be used as a criterion because it is not a property of any single alternative. In addition, it would create dependence among the alternatives which violates the basic assumption of independence of the alternatives and, therefore, is not an admissible criterion. To rate an alternative, one must look at the others to determine whether it is unique or not, and that means that its ranking depends on the other alternatives.

In relative measurement, the effect of the number and relative dominance of the alternatives is naturally captured and implicitly included as a criterion through the normalization process as part of the ratio scale approach. Normalization can be shown to be equivalent to a revision of the weights of the criteria each time the number of alternatives is changed. This is why people's preferences are affected when there is scarcity and abundance and indeed in all situations involving relative comparisons. The rank of the original set of alternatives in the new collection may be different from what the rank was in the absence of new alternatives, and rank reversal or adjustment occurs. Such adjustment, based on what alternatives and how many are considered, is a fait accompli whether one finds it desirable or not. As we said before, some may consider it to be generally desirable (Froman, E.H. 1990) as, for example, when too many copies of a nice hat causes a buyer to choose a slightly less desired one because of the abundance of the first. One apple in a large bowl of fruit may be desirable but many apples is an incentive to choose an orange if it is also potentially desirable. On the other hand one may decide, in a particular application, that rank reversal is inappropriate, and instead gather sufficient experience in advance to use absolute measurement in which the importance of the criteria should not be affected by the number of alternatives considered. For example, universities avidly seek many copies of the very best kind of student. Their value to the university does not depreciate with an increase in their number. But any alternative and not just a copy can influence the rank of the others. That is why the AHP offers the decision maker the ideal or performance mode to preserve rank, and the distributive mode, to allow rank to change. In relative measurement, the ideal mode preserves rank from changing when dominated alternatives are introduced, but not just any alternative that dominates others, partly or completely. In absolute measurement, the ideal mode absolutely preserves rank no matter what kind of alternative is introduced. Since absolute measurement is normative, it parallels what normative utility theorists may have had in mind when they always wanted to preserve rank. There is no reason to absolutely preserve rank even against dominated alternatives in relative measurement, because the quality of an alternative can affect the ranking of the remaining ones. It is not a normative mode.

Multicriteria decisions also deal with intangibles whose magnitudes are unavailable on some known scale and must be estimated by using judgment. This is a fourth reason for using ratio scales. Conflict resolution is replete with intangible factors encountered in negotiations. Ratio scales allow one to pairwise compare tangible and intangible criteria as to their relative contributions to higher criteria or goals. With ratio scales, conflict resolution can be regarded as a tradeoff between gain/loss ratios of the parties each with its respective value system (Saaty, T.L. and Alexander, J.M. 1989). In addition, decision making often involves allocating resources to alternatives in proportion to a ranking based on assessments of ratios of benefits and costs (Saaty, T.L. and Kearns, K.P. 1985), possible with ratio scales but not with interval scales.

In all multicriteria problems, a central issue is how to put together all the dimensions of the problem portrayed by the criteria to evaluate and rank the alternatives. We show below that the use of ratio scales in this process allows us to construct robust additive value functions as approximations to underlying multiplicative functions also measured on ratio scales. We noted earlier that the use of ratio scales makes acceptable the use of additive models, unacceptable with interval scales. In a hierarchy, composition from level to level is equivalent to finding a function f that relates all the attributes on that level to the element in the level below it. Let x_1, \dots, x_n be the attribute values of an alternative, and let y be the composite value of that alternative on all the attributes. Thus,

$$y = f(x_1, \dots, x_n).$$

Let X_1, \dots, X_n be the family of transformations under which the attributes 1, ..., n, respectively, are invariant. Let Y be the corresponding family of transformations for the alternatives (Fechner, G.T. 1860). The functional relationship between the attributes and the alternatives must be the same if $x_i, i=1, 2, \dots, n$, are replaced by $X_i(x_i) = c_i x_i, i=1, 2, \dots, n$, respectively. Thus,

Theorem: If f is continuous in all its arguments, and they are ratio scales, then there exist $\alpha > 0$ and $\beta_i > 0, i=1, 2, \dots, n$, such that:

$$f(x_1, \dots, x_n) = \alpha \prod_{i=1}^n x_i^{\beta_i}.$$

If the arguments are interval scales, then there exist $\alpha > 0$, $\gamma > 0$, and $\beta_i > 0$, $i=1,2,\dots,n$, such that:

$$f(x_1, \dots, x_n) = \alpha \prod_{i=1}^n x_i^{\beta_i} + \gamma.$$

If the arguments are log-interval scales, then there exist $\alpha > 0$, $\gamma > 0$, and $\beta_i > 0$, $i=1,2,\dots,n$, such that:

$$f(x_1, \dots, x_n) = \alpha \prod_{i=1}^n \beta_i \log x_i + \gamma.$$

In the AHP, all the attribute values are estimated in relative terms and normalized to unity. Hence, taking logarithms on both sides of the first expression for $f(x_1, \dots, x_n)$ as a ratio scale, we have:

$$\log f(x_1, \dots, x_n) = \log \alpha + \sum_{i=1}^n \beta_i \log x_i.$$

On using the approximation $\log x \approx x-1$, for $0 < x < 2$, we have the approximate expansion:

$$f(x_1, \dots, x_n) \approx \log \alpha + \sum_{i=1}^n \beta_i x_i.$$

Because normalization implies that $\alpha=1$, we finally have:

$$f(x_1, \dots, x_n) \approx \sum_{i=1}^n \beta_i x_i.$$

thus approximating to a ratio scale in the sense of Stevens (see next section). Note that this approximation can only be met in the case of ratio, but not interval, scales.

Interval and Ratio Scales in the Laboratory

Is there a natural or instinctive justification for interval or ratio scales as a talent or habit of people, or are they forced on us by decision analysts? Two methods are known in psychology for estimating the response to stimuli. The first, magnitude estimation, is called the indirect method by S.S. Stevens (Stevens, S.S. 1946, 1951, 1957, 1971) and is attributed to Fechner and Thurston (Fechner, G.T. 1860; Thurstone, L.L. 1927). The ratios or differences of measurements (on some scale) of a stimulus are formed by making comparisons of two instances of that stimulus.

The second method, more favored by Stevens is called direct measurement. This method is empirical and does not separate the process of detecting and measuring the stimulus from the psychological judgment made in response to it. People assign numerical values to the ratios or differences of stimuli according to their perception. If people are given an arbitrary scale of numbers to represent their responses, it has been observed that they tend to subtract rather than to divide numbers to indicate the

magnitude difference of their sensation when making comparisons. (One notes that people subtract numbers because it is an easier operation and not because they instinctively know it the right thing to do.)

Since the scale chosen to measure the stimulus varies among people, there needs to be agreement on what scale is to be used by groups. When it is not possible to discriminate clearly between the stimuli the measurement obtained is generally inaccurate.

Had people been required to derive a scale for measurement, they would have had no numbers initially to subtract. When forming ratios from comparisons, a natural scale is created for each comparison, where the smaller object is used as the unit and the larger one is estimated as a multiple of that unit.

C.T. Veit has carried out experimental work on the type of scale generated in responding to stimuli (Viet, C.T. 1978). She observed that:

"The present research suggests that judges employ the same subtractive operation whether instructed to estimate the magnitude of simple ratios or to make either category ratings or magnitude estimations of differences. However, judges appear to employ two different operations -- subtractive and ratio -- in the same task when asked to judge ratios of differences. Work done subsequent to this research is in agreement with these conclusions."

Whatever its empirical validity, this observation about using differences more often does not deny that people in fact can use ratios to make comparisons. Other experiments to compare areas, weights and similar physical quantities have shown that people in fact can easily use ratios to make accurate comparisons. In the Analytic Hierarchy Process (AHP) ratio scales are derived from paired comparisons through mathematical operations as distinct from statistical approximations. The eigenvector itself is known to belong to a ratio scale and nothing further is needed to validate it as a ratio scale since

$Acw = cAw = c \lambda_{\max} w$ if and only if $Aw = \lambda_{\max} w$, and when A is nonnegative, w is also nonnegative.

Some psychologists approach ratio scales along Stevens' empirical thinking. Stevens had found that "as a first order approximation" the magnitude of a response can be represented as a power of the magnitude of the stimulus, and hence ratios are needed to compare the stimuli. Accordingly, if the response is a ratio scale it must somehow be a first order approximation to the power of some stimulus. In the absence of a proof that the response belongs to a ratio scale, one relates the response to the stimulus by computing powers through linear regression and determines the power as a ratio of the differences of the logarithms of the response to those of the stimulus. Still, expressing response as a first order approximation in the form of a power of the stimulus measured on a ratio scale does not define a ratio scale. The outcome belongs to a ratio scale only if some very strong conditions are satisfied. Krantz, Luce, Suppes, and Tversky (Krantz, D.H., Luce, R.D., Suppes, P, and Tversky, A. 1971) wrote in referring to the work of S.S. Stevens:

"Surprisingly, he later ... generated an ambiguity in the use of these terms by describing his magnitude estimation scale as a "ratio scale" of measurement. In this experimental procedure, observers are asked to assign numbers to stimuli "in proportion to the sensations evoked," and the resulting numbers are taken to be scale values. In the sense that subjects are asked to produce numbers that preserve subjective "ratios," one sees why this scale might be described as a ratio scale - except for the fact that he earlier introduced the term to refer to those theories in which any two homomorphisms are related by a similarity transformation. Stevens has not provided any argument showing that the procedure of magnitude estimation can be axiomatized so as to result in a ratio-scale representation; he has neither described the empirical relational structure, the numerical relational structure, nor the axioms which permit the

construction of a homomorphism."

These authors give a set of axioms which if empirically satisfied could justify Stevens' claim. Of course if response is derived directly as a ratio scale, it need not be processed through Stevens' power law to prove that it belongs to a ratio scale.

Surreal Numbers and Paired Comparisons (Alling, N.L. 1957; Gonshor, H. 1986; Shulman, P 1995)

A surreal number is a function from an initial segment of the ordinals in the set $\{+,-\}$, an ordinal sequence of pluses and minuses which terminates. If a and b are surreal numbers, $a < b$ if

$a(\alpha) < b(\alpha)$ where α is the first place where a and b differ, with the convention that $(+ -) < (+) < (++)$, which is a linear, lexicographic order (Shulman, P. 1995). Every real number is a rational

in the surreal system. Thus, $\sqrt{2}$ can be written as $\frac{\sqrt{2}\omega}{\omega}$ where ω is the simplest surreal number

larger than all the real numbers. Of interest to the AHP is the paired comparison of two homogenous surreal numbers (like two gods, although unequal and with different potential). Here we note that the existence of an inverse requires the existence of a neutral element which we do not know. The existence of an inverse would allow us to determine how many times one element contains another and how an element can be used as a unit of measurement in paired comparisons. Such an approach would generalize on the AHP.

Conclusion

There are two aspects to the use of scales in multicriteria decisions. The first is the need to represent a problem into a structure that faithfully captures the complexity of this problem. The second is to represent and synthesize with validity the intensity of relations within the structure by means of judgements. Although scales other than ratios may be used to represent intensities, they have a limited scope for arithmetic operations. As a consequence, those scales can only be used for simple models of a problem which may not adequately capture the underlying complexities. Ratio scales, on the other hand, have the needed flexibility in performing arithmetic operations required in modelling a complex decision problem.

References

Allais, M. (1953) Le comportement l'homme rationnel devant le risque: Critique de postulats et axiome de l'ecole Americaine, *Econometrica* 21, 503-546.

Alling, N.L. (1987) *Foundations of Analysis over Surreal Number Fields*, Elsevier Science Publishers, Amsterdam.

Fechner, G.T. (1860) *Elemente der psychophysik*, Leipzig, Germany: Breitkopf and Hartel.

Forman, E.H. (1990) "Facts and Fictions About the Analytic Hierarchy Process," Proceeding of International MCDM meeting, George Mason University.

Gonshor, H. (1986) *An Introduction to the Theory of Surreal Numbers*, Cambridge University Press, Cambridge.

Kahneman, D. and Tversky A. (1979) Prospect theory: An analysis of decision under risk, *Econometrica* 47, 263-291.

- Kahneman, D., Slovic, P., and Tversky, A. (1982) Eds. *Judgment under Uncertainty: Heuristics and Bias*, Cambridge Univ. Press, Cambridge, England.
- Koestler, A. and Smythies, J.R. (Eds.) (1970) *Beyond Reductionism: New Perspectives in the Life of the Sciences*, MacMillan, London.
- Krantz, David H., R. Duncan Luce, Patrick Suppes, & Amos Tversky (1971) *Foundations of Measurement*, Vol. 1 Additive and Polynomial Representations," Academic Press, New York.
- Luce, R.D. and Raiffa, H. (1957) *Games and Decisions*, Wiley, New York.
- MacCrimmon, K.R. (1968) "Descriptive and Normative Implications of the Decision-Theory Postulates," *Risk and Uncertainty* (K. Borch and J. Mosin, Eds.), pp. 3-32, Macmillan Co., New York.
- Saaty, T.L. (1990) *Multicriteria Decision Making : The Analytic Hierarchy Process*, RWS Publications, Pittsburgh.
- Saaty, T.L. and Alexander, J.M. (1989) *Conflict Resolution, The Analytic Hierarchy Approach*, Praeger, New York.
- Saaty, T.L. and Kearns, K.P. (1985) *Analytical Planning: The Organization of Systems*, Pergamon Press, Oxford.
- Schoemaker, P.J.H. (1980) *Experiments on Decisions Under Risk: The Expected Utility Hypothesis*, Martinus Nijhoff, Boston.
- Shulman, P. (1995) "Infinity Plus One, and Other Surreal Numbers," *Discover*, December, 97 - 105.
- Stevens, S. S. (1946) "On the Theory of Scales and Measurement," *Science*, 103, 677-680.
- Stevens, S. S. (1951) "Mathematics, Measurement, & Psychophysics," S.S. Stevens Ed., *Handbook of experimental psychology*, New York: Wiley.
- Stevens, S. S. (1957) "On the Psychophysical Law," *Psychological Review*, 64, 153-181.
- Stevens, S. S. (1971) "Issues in Psychophysical Measurement," *Psychological Review*, 78, 426-450.
- Thurstone, L. L. (1927) "A Law of Comparative Judgment," *Psychological Review*, 34, 273-286.
- Tversky, A. (1969) "Intransitivity of Preference," *Psychological Review*, 76, 31-48.
- Viet, Clarice T. (1978) "Ratio and Subtractive Processes in Psychophysical Judgment," *Journal of Experimental Psychology: General*, Vol. 107, No. 1, 81-107.
- von Winterfeldt, D. and Edwards, W. (1986) *Decision Analysis and Behavioral Research*, Cambridge Univ. Press, Cambridge, England.
- Watson, Stephen, R., and Buede, Dennis M. (1987) *Decision Synthesis: The Principles and Practice of Decision Analysis*, Cambridge Univ. Press, Cambridge, England.