ADDRESSING WITH BREVITY CRITICISMS OF THE ANALYTIC HIERARCHY PROCESS

Thomas L. Saaty
University of Pittsburgh
saaty@katz.pitt.edu

Luis G. Vargas
University of Pittsburgh
lgvargas@pitt.edu

Rozann Whitaker
Creative Decisions Foundation
rozann@creativedecisions.net

A new scientific truth does not triumph by convincing its opponents and making them see the light, but rather because its opponents eventually die, and a new generation grows up that is familiar with it.

-Max Planck

ABSTRACT

The paper provides an overview that covers the main criticisms of the AHP and the authors' replies to them. Because there have been many papers that reply to criticisms, the thrust here is to classify them and reply to them briefly in each category without giving lengthy repetitions of what is already known in the literature.

Keywords: Rank reversal, inconsistent judgments, preserving rank, fundamental scale, pairwise comparisons

http://dx.doi.org/10.13033/ijahp.v1i2.53

1. Introduction

In this paper we address five types of criticisms of the AHP. One is the concern with illegitimate changes in the ranks of the alternatives, called rank reversal, upon changing the structure of the decision. It was believed that rank reversal is legitimate only when criteria or priorities of criteria or changes in judgments are made. Rank reversals were shown by critics to occur when using comparisons and relative measurement that is essential in prioritizing criteria and also alternatives on intangible criteria in two ways: First, when new alternatives are added or old ones deleted; and second, when new criteria are added or old ones deleted with the caveat that the priorities of the alternatives would be tied under these criteria and hence argued that the criteria should be irrelevant when ranking the alternatives. Rank reversals that followed such structural changes were attributed to the use of relative measurement and normalization. Rating alternatives one at a time with respect to the criteria using the ideal mode, always preserves rank. Also, the ideal mode is used with paired comparisons to preserve rank. But rank can and should reverse under more general conditions than had previously been recognized as in introducing copies or near copies of alternatives and criteria turn out not to always be so strictly independent among themselves and from the alternatives. The second concern is about inconsistent judgments and their effect on aggregating such judgments or on deriving priorities from them. A modicum of intransitivity and numerical inconsistency, usually not considered or thought to be permissible in other theories, is permissible in the AHP so that decisions can be treated realistically rather than axiomatically truncated. A condition that may not hold with inconsistent judgments is Pareto optimality. Pareto optimality is an ordinal relation which demands of a method used to aggregate judgments of individuals in a group to a representative collective judgment for that group that when all individuals in the group prefer A to B then the group judgment must prefer A to B. Because in the AHP judgments are not ordinal, it is possible to aggregate the individual judgments into a representative group judgment with or without Pareto optimality. Another condition also inherited from expected utility theory has to do with a relation called Condition of Order Preservation (COP): For all alternatives x_1, x_2, x_3, x_4,such that x_1 dominates x_2 and x_3 dominates x_4, if the evaluator's judgments indicate the extent to which x_1 dominates x_2 is greater than the extent to which x_3 dominates x_4, then the vector of priorities *w* should be such that, not only $w(x_1) > w(x_2)$ and $w(x_3) > w(x_4)$ (preservation of order of preference) but also that $\frac{w(x_1)}{w(x_2)} > \frac{w(x_3)}{w(x_4)}$ (preservation of order of intensity of preference). This condition holds when judgments are consistent but may or may not hold when they are inconsistent. It is axiomatically imposed, sacrificing the original intent of the AHP process to derive priorities that match the reality represented by the judgments without forcing consistency. The third criticism has to do with attempts to preserve rank from irrelevant alternatives by combining the comparison judgments of a single individual using the geometric mean (logarithmic least squares) to derive priorities and also combining the derived priorities on different criteria by using multiplicative weighting synthesis. The fourth criticism has to do with people trying to change the fundamental scale despite the fact that it is theoretically derived and tested by comparing it with numerous other scales on a multiplicity of examples for which the answer was known. The fifth and final criticism has to do with whether or not the pairwise comparisons axioms are behavioral and spontaneous in nature to provide judgments.

Interestingly, the AHP/ANP provides a way to make complex decisions in the most general structures encountered in real life. It makes it possible to derive priorities for all the factors in such structures and synthesize them for an overall outcome, as no other method can because one can build scales for tangibles and intangibles, yet we know little about criticisms of framing and validating problems within such a wide perspective that includes structures, not only for dependence and feedback, but also for benefits, opportunities, costs and risks analyzed separately and then synthesized for the final outcome or in conflict resolution with or without a moderating negotiator.

We give an overview that covers the main criticisms and our replies to them. Because we and others have written numerous papers in reply to criticisms, we have opted to classify them briefly in each category without giving lengthy repetitions of what is already known in the literature.

2. Rank reversal

2.1 Change in structure by adding/deleting alternatives

In relative measurement, unlike measurement on a scale with an arbitrary unit where alternatives are assigned a value independently of other alternatives, when alternatives are compared on several criteria and their weight aggregated, their ranks can change when alternatives are added or deleted (Watson and Freeling 1982; Belton and Gear 1983; Dyer and Ravinder 1983; Dyer 1990). The AHP with its ideal mode preserves rank in rating alternatives (Millet and Saaty 2000). This is equivalent to measuring alternatives one at a time. Adding or deleting alternatives can have no effect on the value and rank of any other alternative. All known software programs that people use implement the ideal mode. In addition when paired comparisons are used, again the ideal mode is often used to preserve rank by idealizing only the first set of alternatives but not after. Thereafter, any new alternative is only compared with the ideal and its priority value is allowed to exceed one before weighting and adding and normalizing. This way the rank of the existing alternatives is always preserved. It is interesting to point out that the distributive mode of the AHP (uniqueness is important), the ideal mode of the AHP (uniqueness is not important), and utility functions (use of interval scales for the ideal), yield the same ranking of alternatives with surprisingly high frequency, except for the case of copies or near copies of an alternative in which the distributive mode always reverses rank, which is legitimate when the uniqueness of the most preferred alternative is important (Saaty and Vargas 1993).

Here is an illustration of rank reversal due to Corbin and Marley (Corbin and Marley 1974).

The first example concerns a lady in a small town, who wishes to buy a hat. She enters the only store in town, and finds two hats, *a* and *b*, that she likes equally well although she leans toward *a*. However, suppose that the sales clerk discovers a third hat, a_1 identically to *a*. Then the lady may well choose hat *b* for sure (rather than risk the possibility of seeing someone wearing a hat just like hers), *a result that contradicts regularity*.

Luce and Raiffa, wrote in their book Games and Decisions, published in 1957 four variations on the axiom about whether rank should or should not be preserved with *counterexamples* in each case and without concluding that it always should and why.

They write:

"Adding new acts to a decision problem under uncertainty, each of which is weakly dominated by or is equivalent to some old act, has no effect on the optimality or non-optimality of an old act.

and elaborate it with

If an act is non optimal for a decision problem under uncertainty, it cannot be made optimal by adding new acts to the problem.

and press it further to

> *The addition of new acts does not transform an old, originally non-optimal act into an optimal one, and it can change an old, originally optimal act into a non-optimal one only if at least one of the new acts is optimal.*

and even go to the extreme with:

> *The addition of new acts to a decision problem under uncertainty never changes old, originally non-optimal acts into optimal ones.*

and finally conclude with:

> *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 question is not whether rank should be preserved, because it is widely believed that it cannot and should not always be preserved (Tversky et al. 1990), but it is whether or not the assumption of independence applies, an assumption used by most multi-criteria methods. *Adding copies or near copies of an alternative until the universe is full of them can depreciate the value and also the rank of that alternative and, as a counter example, negates the possibility of proving a theorem that the rank of independent alternatives must always be preserved when the judgments remain the same and no criteria are added or deleted or their weights changed. A criterion such as "manyness" that represents the number of alternatives cannot be used in the ranking because it forces the dependence of the ranking of each alternative on the existence of every other alternative thus contradicting the assumption of independence. It is illogical* (*or we might say also wrong*) *for all multicriteria methods which all use the rating of alternatives one at a time not to take this into account*. Utilitarian philosophers of the 18th century believed that people ought to desire those things that will maximize their utility. However, this utilitarian viewpoint was abandoned because it was deemed that utility was impossible to measure. Instead, structural accounts of rationality and formal definitions of utility such as rational choice theory were favored. In rational choice theory, the criteria are assumed utility independent and the condition empirically tested. But because the criteria cannot be separated from the alternatives, the resulting weights are not really importance weights but scaling constants. Consequently, according to strong advocates of this theory, independence of the criteria among themselves must be assumed (Keeney and Raiffa 1976; Kamenetzky 1982). Contrary to this assumption, in the AHP/ANP everything can depend on everything else including itself!!! In the AHP/ANP, rank is always allowed to change. It is preserved only when the criteria are conditions imposed on the alternatives and possibly attributes that have had long standing and acquired an importance of their own apart from any particular alternative (Saaty 1991a). For example, we all have the habit of ascribing human kind of rationality to how the universe operates, and assign rationality a high priority. It is not the way some dervishes and ascetics and certainly not the way plants and animals feel about it.

2.2 Change in structure by adding/ deleting criteria

In general, it is known in decision making that if one alters criteria or criteria weights then the outcome of a decision will change possibly leading to rank reversal. This is precisely what some authors use to criticize the AHP. There are two situations. The first is called "wash criteria" which involves the deletion of criteria that are assumed irrelevant because the alternatives have equal or nearly equal priorities under them (Finan and Hurley 2002). The second is called "indifferent criteria" which involves the addition of criteria again assumed irrelevant for the same reason as "wash criteria" (Perez et al. 2006). In the first case the authors made the error of renormalizing the weights of the remaining criteria that then gave rise to rank reversal because the weights of the criteria were changed (Saaty and Vargas 2006). In the second case the addition of a new criterion that was irrelevant also led to rank reversal for exactly the same reason of changing the weights of the criteria. It is surprising that anyone would want to add irrelevant criteria and use it to make an important decision. This approach treats the weights of the criteria not as representative of their importance but as scaling constants like in Multi-Attribute Utility Theory (Keeney and Raiffa 1976).

The correct approach to deal with wash and indifferent criteria is not to delete them or add them but simply to, in the former case, assign zero priorities to the alternatives and keep that criterion, and in the latter case not to add them or if added to, consider this a new decision respecting the influence of added criteria on the final outcome which, as we said above, could lead to different priorities and ranks.

3. Consistency, Pareto optimality and order preservation

Pareto optimality in ordinal preference settings is a condition imposed on preferences which says that if every member of a group prefers A to B, then the group must also prefer A to B. This condition is also known as unanimity. Underlying this condition is the hidden assumption of the transitivity of preferences. In the AHP with its reciprocal condition on the judgments, the geometric mean has been shown to be the unique way to derive a group judgment from the individual judgments under fairly general conditions. Note that Pareto optimality as used in economic and social practice applies to a final ordering of each individual of all the alternatives and not to judgments that obtain that order. In the AHP because preference order is indicated by priorities rather than by an ordinal statement of preference, Pareto optimality always holds when the stated condition is satisfied, and there is no problem with Pareto optimality.

When Pareto optimality is applied to judgments, there two possibilities: The first is when all judgments in a pairwise comparison matrix $A=(a_{ij})$ are consistent (i.e., $a_{ij}a_{jk}=a_{ik}, \forall i, j, k$ and a_{ij} have the form $a_{ij}=w_i/w_j$ where the w_i 's are the priorities), in which case one has transitivity and also Pareto optimality. The second is when the judgments are inconsistent. In this case Pareto optimality holds under restrictive conditions like row dominance for each individual, i.e., there is an ordering of the rows and corresponding judgments in each row.

One may ask: Why should Pareto optimality be imposed on a method that uses cardinal preferences when it already has a process for aggregating individual judgments, along with the importance of the individuals involved, into a group judgment? If the members of the group are agreeable to using the geometric mean to combine their judgments, even if Pareto optimality is not satisfied, why should their combined judgment be any less valid than any other procedure that satisfies Pareto optimality?

Finally, Pareto optimality is not universally regarded as a desired condition in all decisions. A common criticism of a state of Pareto efficiency is that it does not necessarily result in a socially desirable distribution of resources, as it may lead to unjust and inefficient inequities (Sen 1993; Barr 2004).

A condition that mirrors preferences expressed with interval scale value functions is the Condition of Order Preservation (COP) (Bana e Costa and Vansnick 2008). In interval scale value theory, a value function *v* must satisfy the condition that if a consequence *i* is preferred to a consequence *j* more than a consequence *h* is preferred to a consequence *k* then v(*i*)-v(*j*)>v(*h*)-v(*k*). Note that preferences are ordinal and hence no intensity of preference or judgment is involved. On the other hand, an individual imposing COP assigns judgments to the preferences. Thus, if $a_{ij} > a_{hk}$ then $\frac{w_i}{w_j} > \frac{w_h}{w_k}$. This condition is always satisfied if the judgments are consistent because all logical methods of deriving priorities yield the same priorities. When the judgments are inconsistent, only the eigenvector obtains priorities that capture the transitivity of dominance reflected in the judgments. A major property of consistent judgments arranged in a matrix $A = (a_{ij})$ is that it satisfies the condition $A^k = n^{k-1}A$, where n is the order of *A*, so all powers of *A* are essentially equal to *A*. Now dominance of an inconsistent matrix no longer satisfies this condition and one must consider priorities derived from direct dominance as in the matrix itself, second order dominance obtained from the square of the matrix and so on. The total dominance of each element is obtained as the normalized sum of its rows. The result is an infinite number of priority vectors each representing a different order of dominance. The Cesaro sum of these vectors is equal to the priority vector obtained from the limiting powers of the matrix. Thus, only the eigenvector gives the correct ordering and priority values. COP imposes a condition on the priorities based solely on the original preferences without regard to dominance of higher order, and it thus likely to lead to the wrong priorities and order. In fact, we know of the existence of examples to support this statement (Salomon 2008). COP was devised for use in a method known as MACBETH (Bana e Costa et al. 2003). However, the value functions derived are interval scales, so COP is expressed as ratios of differences. Finding the value function that satisfies COP in MACBETH involves an optimization technique that yields a non-unique solution!!

To summarize, COP forces the condition that $a_{ij} > 1$ should imply $w_i > w_j$, which is not always true when the judgments are inconsistent; violates the integrity of the eigenvector as the way to derive priorities capturing higher order interactions among the judgments; artificially forces adjustment of the judgments without asking the decision maker if the altered value is acceptable in his framework of understanding; and yields invalid results for single matrices with known measurements.

4. Priority derivation and synthesis with the geometric mean

A number of people, in their concern with always preserving rank, look for schemes to synthesize inconsistent judgments and also priorities (Holder 1990). The only other method that has been proposed and pursued in the literature has been the geometric mean for a single matrix (Barzilai 1997), in which the elements in each row of the matrix are multiplied, the nth root taken and the resulting vector normalized. This process does not capture the effect of transitivity of dominance in the case of inconsistent judgments and hence, it can lead to wrong priorities and order (Saaty 1991b).

Synthesizing priorities derived in any manner by raising them to the power of the priority of the corresponding criterion and then multiplying the outcome (Lootsma 1993; Barzilai and Lootsma 1997) has the shortcoming that $0 < x < y < 1$ and $0 < p < q \Rightarrow x^p > y^q$ for some p and q. This means that an alternative that has a smaller value under a less important criterion is considered to be more important than an alternative that has a larger value under a more important criterion, which is absurd.

One can also show the absurdity of this process of synthesis because it yields wrong known results. By considering alternatives with known measurements under two or more criteria which then inherit their importance from the measurements under them, normalizing these measurements, raising them to the power of the priority of their corresponding criterion and multiplying, one obtains a different outcome than simply adding the measurements and then normalizing them (Vargas 1997).

5. Altering the Fundamental Scale

A number of authors have proposed changes in the fundamental 1-9 scale of the AHP more as a passing suggesting without either a proof of the resulting improvement if any or validation examples to test their assertions (Ma and Zheng 1991; Salo and Hamalainen 1997).

6. Are the axioms about comparisons behaviorally meaningful?

People who subscribe to expected utility theory claim (see for example, Dyer 1990, p.251) "…each of these axioms has a clear and obvious meaning as a description of choice behavior. Therefore, each axiom can be debated on the basis of its appeal as a normative descriptor of rationality, and each axiom can also be subjected to empirical testing." This statement is the basis for the criticism of the fundamental scale in the AHP. Are paired comparisons behaviorally based or are they an invention of ours? The Harvard psychologist Arthur Blumenthal (Blumenthal 1977) believes that there are two types of judgment: "*Comparative judgment* which is the identification of some relation between two stimuli both present to the observer, and *absolute judgment* which involves the relation between a single stimulus and some information held in short term memory about some former comparison stimuli or about some previously experienced measurement scale using which the observer rates the single stimulus." In the Analytic Hierarchy Process (AHP) we call the first *relative measurement* and the second *absolute measurement*. In relative measurement we compare each alternative with many other alternatives and in absolute measurement we compare each alternative with one ideal alternative we know of or can imagine, a process we call *rating alternatives*. The first is descriptive and is conditioned by our observational ability and experience and the second is normative, conditioned by what we know is best, which of course is relative. Comparisons must precede ratings because ideals can only be created through experience using comparisons to arrive at what seems best. It is interesting that in order to rate alternatives with respect to an ideal as if they are independent can only be done after having made comparisons that involve dependence to create the ideal or standard in the first place. Making comparisons is fundamental and intrinsic in us. They are not an intellectual invention nor are they something that can be ignored.

The need for quantifying the intensity of preferences is all around us. Donald J. Boudreaux writes:
"My third reason for not voting is that voting registers only each voter's order of preferences and not that voter's intensity of preferences. Unlike in private markets where I can refuse to buy a good or service if I judge its price to be too high—and then decide to buy that same product if its price falls—in elections each voter merely gets to say which candidate he prefers above all who are on the ballot. If I vote for Smith rather than Jones, this means only that I prefer Smith to Jones. My vote for Smith reveals nothing about how much I prefer Smith to Jones." (Boudreaux 2008)

Paired comparisons consist of two steps. First, as in utility theory, there is a binary comparison, for example, alternative A is preferred to alternative B. Second, we must decide with how much more intensity we prefer A to B. Because in expected utility theory preferences are built on lotteries, it is already assumed that intensity of preference is accounted for, even though utilities do not always represent intensity of preference (Sarin 1982). Without the probability function one is left with ordinal utilities which yield only ranking. Probabilities play the role of the fundamental scale in the AHP. On the other hand, the AHP articulates the intensity of pairwise comparison preferences using an instinctively built-in absolute scale. The mathematician and cognitive neuropsychologist Stanislas Dehaene writes: "Introspection suggests that we can mentally represent the meaning of numbers 1 through 9 with actual acuity. Indeed, these symbols seem equivalent to us. They all seem equally easy to work with, and we feel that we can add or compare any two digits in a small and fixed amount of time like a computer." (Dehaene 1997)

Pareto (1848-1923) rejected altogether the idea that quantities of utility mattered. He observed that if we map preferences onto Edgeworth's indifference curves, we know everything necessary for economic analysis. To map these preferences, we make *pairwise comparisons* between possible consumption bundles. The agent will either be indifferent between each bundle, or else will prefer one to the other. By obtaining comparisons between all bundles, we can draw a complete map of an individual's utility. These comparisons were ordinal in nature and did not go far enough to represent intensity of preference.

7. General observations

We gave above arguments about the major issues. The references include many papers we know about, our published responses to some and also references to papers we wrote mostly on the subject of rank preservation and reversal.

The first paper questioning some aspect of the AHP was that by Watson and Freeling (Watson and Freeling 1982). The authors questioned the validity of the questioning process by means of which judgments are elicited (Saaty et al. 1983). Belton and Gear (Belton and Gear 1983) built an example of a simple hierarchy with three criteria and 3 alternatives, and showed that adding a copy of an alternative made rank reversal was possible (Saaty and Vargas 1984). The same problem was reported (Dyer and Ravinder 1983). Later Dyer (Dyer 1990) used the same arguments to challenge the validity of the axioms and the principle of hierarchic composition, and provided his own solution which he considered to be consistent with expected utility theory!! (Harker and Vargas 1990; Saaty 1990)

Holder (Holder 1990) criticized the eigenvector method by questioning the validity of the optics experiment and the principle of hierarchic composition, for the same reason which was rank reversal (Saaty 1991b). The same criticisms were voiced in (Lootsma 1993; Salo and Hamalainen 1997; Finan and Hurley 2002; Hurley 2002; Perez et al. 2006). All these authors criticize the principle of hierarchy composition. Salo and Hamalainen (Salo and Hamalainen 1997) also criticize the composition principle in the Analytic Network process.

Other authors have criticized the AHP on the grounds that the 1-9 scale is not appropriate (Ma and Zheng 1991; Lootsma 1993; Salo and Hamalainen 1997).

In group decision making the geometric mean has been criticized because it violates Pareto optimality (Lootsma 1993).

There have been people who expect to put their own default numbers in an AHP structure without input about the particular decision and get rational numerical outcomes. One such person has published notes against the AHP and other decision methods with strongly made arguments mostly published in unrefereed journals is Jonathan Barzilai (Barzilai 1998). He has been promising for many years to provide the scientific community his own decision theory. The third author has shown (Whitaker 2004; Whitaker 2007a) in detail where his thinking is in error. One of his fundamental assumptions is that in order for paired comparisons to be valid, the underlying scale must be a ratio scale. He totally ignores the fact that paired comparison judgments are represented by numbers from an absolute scale and that the derived priority scales are relative scales of absolute numbers with no zero and no unit. In addition, he assumes that criteria are equally weighted when in fact he has assigned actual measurements to the alternatives on a known scale. It can be easily shown that when several criteria have alternatives measured on the same scale the weights of the criteria are given by the sum of the measurements of the alternatives under each divided by the sum under all the criteria. In that case, weighting the relative measurements of the alternatives under each criterion by the weight of that criterion and adding over the criteria yields the same relative measurement as one obtains by first simply summing the measurements for each alternative and then normalizing the final sums for all the alternatives to put their total measurements in relative form. This is a common error that people make. Indeed it is the same error made by S. Schenkerman (Schenkerman 1997) in his paper in which he starts out by assuming some measurements for the alternatives with respect to each criterion so that he can perform arithmetic operations on them to get an answer. He then sets up an AHP problem to see if it gives the same answer as he got, but in relative terms. He assigns the criteria equal weights, but he has already given the alternatives measurements under each criterion and a fortiori he is no longer free to assign the criteria arbitrary weights of his own choice. Here we note that there are many ways to combine measurements on objects using formulas to get an overall answer. It is naïve to assume that the straightforward process of weighting and adding in a hierarchic structure should always yield results that coincide with the formula even if one can, by carefully modeling the problem as a decision, use the hierarchic weighting approach to obtain the same answer.

For attributes/properties for which a scale has not yet been developed, Barzilai assumes that there cannot be information about them that can be measured and hence paired comparisons with respect to criteria are invalid. He announces by fiat and without proof that hierarchic composition is linear and that it generates nonequivalent value functions from equivalent decompositions. In fact both theory and many examples show that hierarchic composition is nonlinear, and the value functions generated are valid when it is done correctly.

Replies to the issues in such papers have been properly addressed in the literature and will not be repeated here.

8. Conclusions – our concern with validation in decision making in general

It is considered scientifically justifiable to require some sort of objective validation of numbers derived as answers in decision making. People in the field of decision making, particularly the normative kind, seem to be oblivious to the issue of validation as if it is a requirement they do not have to heed. It is true that judgments and priorities are subjective, but this does not mean that what a decision maker obtains by following the number crunching dictates of some theory will be justifiable to use in practice. It may be that results from their theory appear reasonable to the creators of it who are conditioned by a few techniques they know well, but they may have no real credibility in practice. Nor is the consent of the decision maker proof of anything because he may not be sophisticated in demanding justification according to more rigid standards of knowledge and practice. Nor is it proof that the technique is right if the decision outcome worked out successfully one time or even a few times.

The AHP is a psychophysical theory that finds some of its validations in measurement itself. Here are two examples and there are many others that would fill a book. Some are with single matrices, some with hierarchies and some even with networks (Whitaker 2007b). For brevity and to give the reader an idea of how it is done, we illustrate with two simple examples here.

An audience of about 30 people, using the AHP 1-9 scale with reciprocal values and coming to consensus on each judgment (instead of the geometric mean which is the proven way to use to combine judgments in the AHP), provided judgments from their general knowledge and experience about what people drink to estimate the *dominance* of the consumption of drinks in the United States (which drink listed on the left of Table 1 is consumed more in the US over a drink listed at the top of Table 1, and how much more than that drink?). The derived vector of relative consumption and the actual vector, obtained by normalizing the consumption given in official statistical data sources, are at the bottom of Table 1.

Table 1 Which drink is consumed more in the U.S.?

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

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

Coffee	Wine	Tea	Beer	Sodas	Milk	Water
.177	.019	.042	.116	.190	.129	.327

with a consistency ratio of .022.

The actual consumption (from statistical sources) is:

.180	.010	.040	.120	.180	.140	.330

Those who did the example could not possibly have known the answers in advance but the results confirmed the accuracy of their judgments.

Recently the second author applied judgments to estimate the relative size of the populations of seven cities in Spain. The judgments and the close outcome when the priorities are compared to the relative actual values are shown in Table 2.

Table 2 Which city has the larger population?

	Madrid	Barcelona	Valencia	Sevilla	Zaragoza	Malaga	Bilbao	Priorities	Actual in millions	Relative actual
Madrid	1	2	5	5	6	6	9	0.429	3.400.000	0.434
Barcelona	1/2	1	2	2	3	3	4	0.197	1.500.000	0.192
Valencia	1/5	1/2	1	1	1	1.5	2	0.091	740.000	0.095
Sevilla	1/5	1/2	1	1	1	1	2	0.086	700.000	0.089
Zaragoza	1/6	1/3	1	1	1	1	2	0.079	600.000	0.077
Malaga	1/6	1/3	1/1.5	1	1	1	1	0.068	528.000	0.067
Bilbao	1/9	1/4	1/2	1/2	1/2	1	1	0.048	358.000	0.046

We recommend that multicriteria methods put greater emphasis on validation to acquire greater credibility in practice. Validation is much more difficult when all judgments depend on feelings alone without memory from the senses, and when the criteria are all intangible. But there are other ways to improve the credibility of the outcome that have been discussed in the literature (Whitaker 2007b).

REFERENCES

Bana e Costa, C.A., J.-M. D. Corte and J.-C. Vansnick (2003). MACBETH. Working Paper 03.56. London, London School of Economics.

Bana e Costa, C.A. and J.-C. Vansnick (2008). A critical analysis of the eigenvalue method used to derive priorities in the AHP. *European Journal of Operational Research* **187**(3) 1422-1428.

Barr, N. (2004). *Economics of the Welfare State*. New York, Oxford University Press (USA).

Barzilai, J. (1997). Deriving weights from pairwise comparison matrices. *Journal of the Operational Research Society* **48**(12) 1226-1232.

Barzilai, J. (1998). On the Decomposition of Value Functions. *Operations Research Letters* **22** 159-170.

Barzilai, J. and F. A. Lootsma (1997). Power Relations and Group Aggregation in the Multiplicative AHP and SMART. *Journal of Multi-Criteria Decision Analysis* **6** 155-165.

Belton, V. & A.E. Gear (1983). On a Short-coming of Saaty's Method of Analytic Hierarchies. *Omega* **11**(3) 228-230.

Blumenthal, A. (1977). *The Process of Cognition*. Englewood Cliffs, New Jersey, Prentice-Hall, Inc.

Boudreaux, D. J. (2008). The Freeman: Ideas on Liberty *Foundation for Economic Education* **58**(3).

Corbin, R. & A.A.J. Marley (1974). Random Utility Models with Equality: An Apparent, but not Actual, Generalization of Random Utility Models. *Journal of Mathematical Psychology* **11** 274-293.

Dehaene, S. (1997). *The Number Sense*, Oxford University Press.

Dyer, J.S. (1990). Remarks on The Analytic Hierarchy Process. *Management Science* **36**(3) 249-258.

Dyer, J.S. & H.V. Ravinder (1983). Irrelevant Alternatives and the Analytic Hierarchy Process. Working Paper, The University of Texas at Austin.

Finan a& Hurley (2002). The Analytic Hierarchy Process: Can Wash Criteria Be Ignored? *Computers and Operations Research* **29**(8) 1025-1030.

Harker, P.T. & L.G. Vargas (1990). Reply to "Remarks on The Analytic Hierarchy Process" By J.S. Dyer. *Management Science* **36**(3) 269-273.

Holder, R.D. (1990). Some Comments on the Analytic Hierarchy Process. *Journal of the Operational Research Society* **41**(11) 1073-1076.

Hurley, W.J. (2002). Letters to the Editor: Strategic Risk Assessment. *Canadian Military Journal* **Summer** 3-4.

Kamenetzky, R.D. (1982). The Relationship Between the Analytic Hierarchy Process and the Additive Value Function. *Decision Sciences* **13**(4) 702-713.

Keeney, R.L. & H. Raiffa (1976). *Decisions with Multiple Objectives: Preferences and value Tradeoffs*. New York: John Wiley & Sons.

Lootsma, F.A. (1993). Scale Sensitivity in the Multiplicative AHP and SMART. *Journal of Multi-Criteria Decision Analysis* **2,** 87-110.

Ma, D. &X. Zheng (1991). 9/9-9/1 Scale Method of AHP. Proceedings of the 2nd Int.'l Symposium on the AHP, Pittsburgh, PA, University of Pittsburgh. 1, 197-202.

Millet, I. & T.L. Saaty (2000). On the relativity of relative measures – accommodating both rank preservation and rank reversals in the AHP. *European Journal of Operational Research* **121,** 205-212.

Perez, J., J.L. Jimeno & E. Mokotoff (2006). Another Potential Shortcoming of AHP. *TOP* **14**(1) 99-111.

Saaty, T.L. (1990). An Exposition of the AHP in Reply to the Paper: Remarks on the Analytic Hierarchy Process. *Management Science* **36**(3) 259-268.

Saaty, T.L. (1991a). Rank and the Controversy About the Axioms of Utility Theory ─ A Comparison of AHP and MAUT. The 2nd International Symposium on The Analytic Hierarchy Process, Pittsburgh, PA.

Saaty, T.L. (1991b). Response to Holder's Comments on the Analytic Hierarchy Process. *The Journal of the Operational Research Society* **42**(10) 909-914.

Saaty, T. L. & L.G. Vargas (1984). The Legitimacy of Rank Reversal. *Omega* **12**(5) 513-516.

Saaty, T.L. & L.G. Vargas (1993). Experiments on Rank Preservation and Reversal in Relative Measurement. *Mathematical and Computer Modeling* **17**(4/5) 13-18.

Saaty, T.L. & L.G. Vargas (2006). The analytic hierarchy process: wash criteria should not be ignored. *Int'l J'l of Management and Decision Making* **7**(2/3) 180-188.

Saaty, T.L., L.G. Vargas & R. E. Wendell (1983). Assessing Attribute Weights by Ratios. *Omega* **11**(1) 9-13.

Salo, A.A. & R.P. Hamalainen (1997). On the Measurement of Preferences in the Analytic Hierarchy Process. *Journal of Multi-Criteria Decision Analysis* **6**(6) 309-319.

Salomon, V.A O. (2008). An Example of the Unreliability of MACBETH Applications. 4th International Conference on Production Research, June, Sao Paulo, Brazil.

Sarin, R.K. (1982). Strength of Preference and Risky Choice. *Operations Research* **30**(5) 982-997.

Sen, A. (1993). Markets and Freedom: Achievements and limitations of the market mechanism in promoting individual freedoms. *Oxford Economic Papers* **45**(4) 519-541.

Schenkerman, S. (1997). Inducement of Nonexistent Order by the Analytic Hierarchy Process. *Decision Sciences* **28** (2) 475-482.

Tversky, A., P. Slovic & D. Kahneman (1990). The Causes of Preference Reversal. *The American Economic Review* **80**(1) 204-215.

Vargas, L.G. (1997). Why the Multiplicative AHP is Invalid: A practical counterexample. *Journal of Multi-Criteria Decision Analysis* **6,** 169-170.

Watson, S.R. & A.N.S. Freeling (1982). Assessing Attribute Weights. *Omega* **10**(6) 582-583.

Whitaker, R. (2004). Why Barzilai's Criticisms of the AHP are Incorrect. Int'l Meeting of the Multi-Criteria Decision Making Society, Whistler, Canada.

Whitaker, R. (2007a). Criticisms of the Analytic Hierarchy Process: Why they often make no sense. *Mathematical and Computer Modeling* **46**(7/8) 948-961.

Whitaker, R. (2007b). Validation Examples of the Analytic Hierarchy Process and Analytic Network Process. *Mathematical and Computer Modelling* **46**(7/8) 840-859.