AN INVESTIGATION OF THE AHP JUDGEMENT RENDERING PROCESS

EDWARD J. LUSK
The Wharton School
University of Pennsylvania
TRI-NEB Building/S2
Philadelphia, PA 19104
U.S.A.

JOAO NEVES
Catholic University
Lisboa
Portugal

MADJID TAVANA
Beaver College
Glenside, PA
U.S.A.

ABSTRACT

The Analytic Hierarchy Process (AHP) is discussed and illustrated. A controlled experiment is conducted to ascertain if the quality of judgements rendered using the AHP is task dependent or is affected by psychological or demographic variables. No evidence is found to suggest that sex, disembedding skill or measures of verbal or mathematical scholastic apptitude are useful in predicting the ability to render consistent judgements using the analytic hierarchy process. Further, the performance across tasks does not vary systematically. Therefore, in organizing groups or individuals to render judgements using the AHP, sex, disembedding skill, scholastic apptitude measured by the SAT, or pervious performance rendering judgements using the AHP should not be used as screening criteria.

1. INTRODUCTION

Decision makers are often required to establish priorities among alternatives. A technique which has been used in a variety of situations

Journal of Organizational Behaviour and Statistics
Vol. 1, Nos. 2, 3, 4 ; pp. 133—147

to establish priority weights for a set of alternatives is the Analytic Hierarchy Process (AHP) developed by Saaty (1977); also see Saaty (1980). Using the AHP, the decision maker renders pairwise preference judgements which are formulated into priority weights. Such priority weights are presumed to guide the decision making process. Since rendering judgements using the AHP is likely to be a demanding cognitive task, two questions of interest arise in these judgement situations : 1) do psychological or demographic variables affect the quality of the judgements rendered, and 2) is performance in one judgement situation associated with judgemental performance in a different situation?

How psychological and demographic variables affect judgements ellicited using the AHP is particularly important in those situations where a number of individuals render judgements and an overall priority vector is to be synthesized from the individual judgements. If the quality of the individual judgements varies in a systematic way or if judgemental performance is highly associated across judgement tasks, such information would also be useful in organizing individuals to cope with specific decision problems using the analytic hierarchy process.

The purposes of this paper are to 1) present the results of a controlled experiment which examines the relationship between psychological and demographic factors and the quality of the judgement process using the AHP, 2) examine the quality of individual judgements across judgement tasks, and 3) discuss the implications of these research findings regarding setting priorities using the AHP. Consider now a discussion of certain technical aspects of the AHP.

2. THE ANALYTIC HIERARCHY PROCESS

The AHP model uses a measurement scale which has only the following elements and their reciprocals :

$$[1, 2, 3, 4, 5, 6, 7, 8, 9].$$

The numeric descriptors of this scale are taken to mean :

1 if alternative a^u and alternative a^v are equally important on a particular criterion,

3 if a^u is weakly more important then a^v on a particular criterion,

5 if a^x is strongly more important then a^y on a particular criterion,

7 if x^a is demonstrably more important than a^y on a particular criterion,

9 if a^a is aboslutely more important than a^y on a particular criterion.

Intermediate values 2, 4, 6 and 8 may be used if the primary descriptors (1, 3, 5, 7, 9) are not adequate to represent an individual's judgements.

The result of the pairwise evaluation of two alternatives in terms of a criteria will be noted by specifying one of the scale descriptors and entering that value in the appropriate row and column of a judgement matrix for a given criteria as follows :

$$[A]\, j = a_j^{kh}, \quad h, k = 1, \ldots, n, j = 1, \ldots, m,$$

where : the superscripts k and h indicate that the kth alternative is compared to the hth alternative and the result of that comparison is quantified using the above scale system and recorded in the kth row and hth column of A, and j denotes a particular criterion, $1 \leqslant j \leqslant m$ for any a^{kh}. If $k=h$, then the value $a^{kh}=1$ for any alternative. Finally, the following reciprocal relationship is assumed : $a^{kh}=1/a^{hk}$.

After the m judgement matrices are completed, the eigenvector associated with the largest eigenvalue is abstracted for each judgement matrix and organized by columns into an $n \times m$ matrix. Whenever more than one criteria is involved, a criteria matrix must be formed in a manner similar to that of the alternative matrices. Then the eigenvector associated with the criteria matrix is calculated, note this eigenvector as c. The final priority matrix is then formed as :

$$\underset{n \times m}{N \cdot [(w_1, w_2, \ldots, w_m)} \cdot \underset{m \times 1}{(c)}] = \underset{n \times 1}{p}$$

where : N is the norm operation which causes the n elements of i to sum to one,

: wj is the eigenvector associated with the judgement matrix developed from jth criteria,

: c is the eigenvector associated with the criteria matrix,

: p is the final priority vector for the n alternatives.

The way a judgement matrix is constructed also allows the computation of an eigenvalue, noted as λ_{max}, which is directly related to the relative intransitivity of the entries of a judgement matrix. For example, assume that criteria x, y and z are to be evaluated respecting the pairwise scaling system. If x contrast y is given a 3, and y contrast z is given a 3, and x contrast z is given a 9, then all the pairwise comparisons in that 3×3 matrix are transitive. In this case, the eigenvalue of matrix A would equal 3 which indicates that the judgements were transitive. If the judgements were not transitive, the value of the eigenvalue would be greater than the trace of the judgement matrix, 3 in this case. The greater the intransitivity the greater, would be the value of the eigenvalue until it reaches its maximum. For the case of a 3×3 matrix, this value is 91/9.

The following two measures of judgement matrix inconsistency have been proposed by Saaty (1977) and Vargas (1980) respectively :

$$\sum_{ijk} a_{ij}\, a_{jk} - a_{ik}, \text{ and}$$

$$1/(n(n-1)(n-2)) \sum_{ijk=1} a_{ij}\, a_{jk}\, a_{ki} .$$

Both measures are reasonably well surrogated for by the value of the eigenvalue of the judgement matrix (Vargas (1980) and Lusk (1979)). For this reason, the measure of inconsistency to be used in assessing the quality of the judgement process will be the following statistic suggested by Saaty (1977) :

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

Our interest is in the ability to render consistent judgements rather than on how the level of inconsistency affects the final priority vector, i. Therefore, our study will concentrate on the rendering of judgements for a set of alternatives for one criterion. This facilitates the design of the experiment ; further there are numerous applications which require the evaluation of criteria relative to a higher order criterion. For example, a group of physicians and managers who owned a chain of hospitals were dissatisfied with the way capital investments were evaluated. They were concerned because after the capital budget was approved, it was difficult to adequately explain why certain projects were approved while other

projects were not approved. According to the decision makers, the projects seemed to be evaluated and approved randomly or on some unknown basis. Management believed that the major problem was the absence of clear criteria for differentiating among investement alternatives. The president of the organization was concerned because in the past this and similar groups never were able to develop and agree upon a way to evaluate investments.

After some preliminary discussion, the group formulated the following criteria for differentiating among investments :

C_1 : Improve Patient Welfare (PW)

C_2 : Provide for Maximization of Long-run Profitability (LRP), and

C_3 : Improve the Status of the Hospital Group (S).

These three criteria were to be evaluated with respect to their importance in facilitating the achievement of the mission of the organization (which at this point was not specified explicitly). After about two hour of discussion, the group arrived at the following judgement matrix.

$$A\,(3\times 3)=\begin{array}{c|ccc} & PW & LRP & S \\ PW & 1 & 1/5 & 3 \\ LRP & 5 & 1 & 1/5 \\ S & 1/3 & 5 & 1 \end{array}$$

Based upon the inconsistency measure derived from this judgement matrix, ($\mu=1.2$), the group decided that their judgements were too inconsistent to warrant further consideration. In this case, the judgement matrix recorded that long-run profitability was more important than patient welfare and patient welfare was more important than improved status which implied that long-run profitability was more important than improved status. However, the converse is recorded in the judgement matrix, since the LRP, S contrast is given a 1/5. As the president of the hospital group remarked "The chance that our group process resulted in a logical resolution of the relationship among patient welfare, long-run profitability, and status is too low. We' ve missed somewhere along the way". The group then began the process of developing a clear statement of the mission which could be used to develop a better sense of the relationship between C_1, C_2 and C_3. This example is used to illustrate how intransi-

tivity information may be used to monitor the quality of the judgement process. The question addressed by the research is : if individuals develop an intransitive matrix, are they likely to develop an intransitive matrix given their next priority setting task, or is there a common psychological or demographic variable which acts to cause intransitive judgements. Consider now the psychological test selection and experimental design.

3 PSYCHOLOGICAL TEST SELECTION AND DEMOGRAPHIC INFORMATION

Developing a judgement matrix requires the assessment of the relative intensity of alternatives on the measurement scale. The AHP method does not require that the person developing the judgements consider the judgement information recorded previously so that the pair of alternatives currently under consideration are consistent with the previously recorded information. Therefore, AHP permits the recording of inconsistent information. Recording pairwise judgements so they are consistent is likely to be a complicated cognitive task requiring the recognition of the dimensionality of the associations and how that dimensionality is reweighted as the pairs of alternatives change.

After considering a number of psychological tests, the Group Embedded Figures Test (GEFT) was selected as a measure which may be associated with the ability to develop consistent judgementma trices. The GEFT requires the abstraction of a simple figure from a complex background. Such a task embodies both cognitive and psychomotor skills to identify and locate the simple figure in the comp'ex backgfound. Previous work on the GEFT suggests that individuals who score relatively high on the test are better at tasks requiring the organization and relative valuation of specific information. Witkin and his associates (1967) suggest :

The extent of differentiation is reflected in the area of perception in the degree of field dependence or independence. In a field-dependent mode of perceiving, perception is dominated by the overall organization of the field ; there is relative inability to perceive parts of a field of discrete. This global quality is indicative of limited differentiation. Conversely, a field independent style of perceiving, in which parts of a field are experienced as discrete from organized background, rather than fused with it, is a relatively differentiated way of functioning.

Even though the GEFT is a perceptual test, Witkin indicates that the ability to disembed a simple figure from a complex background transcends that simple task and manifests itself, in problem-solving situations as well. Witkin et al. (1962, 1971) note :

The significant relations frequently reported between measures of field dependence/independence and total standard intelligence test scores is carried largely by these portions of intelligence tests which require analytical functioning. In other words, the relation is based on the expression of a particular style of field approach to both.

The finding of a linkage between analytical and structuring abilities suggested that the cognitive style involved was even broader than implied by the "global-analytical" concept ... The cognitive-style concept approaches perceptual and intellectual activities from the perspective of the person engaged in them. Pursuit of this approach has demonstrated that an individual shows the same characteristic ways of functioning across these activities, suggesting that the classical division between the perceptua and the intellectual needs to be relaxed.

Results consistent with Witkin's assertions that the GEFT measures not only perceptual skills but may be used to infer the nature of the process by which information is developed may be found in Hess (1966), Huysmans (1970), Bariff and Lusk (1977), Doktor and Hamilton (1973) and Benbasat and Dexter (1979). For example, Benbasat and Dexter note their results indicate that high analytic individuals use a relatively structured approach to solving problems, whereas the low analytical individuals use a relatively unstructured approach.

Considering this information, it seems reasonable to expect that if the judgement rendering process taps into the same skills as required to perform simple abstraction, the GEFT could be expected to be associated with that availability. Such an association would facilitate the cognitive dimensions underlying the judgement rendering process.

In addition to the GEFT information, sex, and SAT scores, will be collected to ascertain if these factors are important in accounting for ifferences in consistency scores.

4. EXPERIMENTAL DESIGN

To ascertain if an individual's ability to render consistent judgement is affected by the particular task or is affected by personality or demographic variables, the following three different judgement problems situations were developed and pretested.

Case 1 : Special Proximities

The subjects, who were juniors registered in a production course, were given the graphic and task represented in Figure 1 and the following judgement scale :

Record your judgement using the following scale :

Judgement Value	Definition	Explanation
1	Equally distant	City *i* and city *j* are *equally distant* from city P.
3	Slightly further	City *i* is *slightly further* from city P than city *j* is from city P.
5	Significantly further	City *i* is *significantly further* from city P than city *j* is from city P.
7	Much further	City *i* is *much further* from city P than city *j* is from city P.
9	Overwhelmingly further	City *i* is *overwhelmingly further* from city P than city *j* is from city P.

2, 4, 6, 8 Intermediate values between the two adjacent judgement values.

Reciprocals. If the city on the left is *closer* to P than is the city on the top, use the reciprocals of the scale value. For example, if the city *j* (on the top) is much further from city P than the city *i* (on the left), one should enter 1/7 in cell (*i*, *j*).

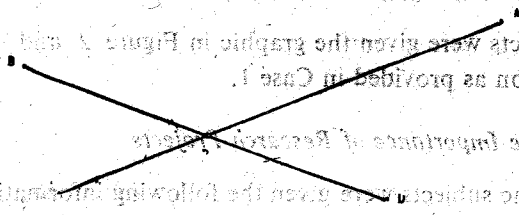

Figure 1.
Relative Distances to Specific Points : Case 1

A, B, C and D are four cities whose relative distances from city P are to be estimated.

Note : before making any judgments please read the entire page.

	J	A	B	C	D
A		1			
B			1		
C				1	
D					

Given a city (i) on the left side of the above table and another on the top (j) *how much further* is the first city from P than the second city is from P ?

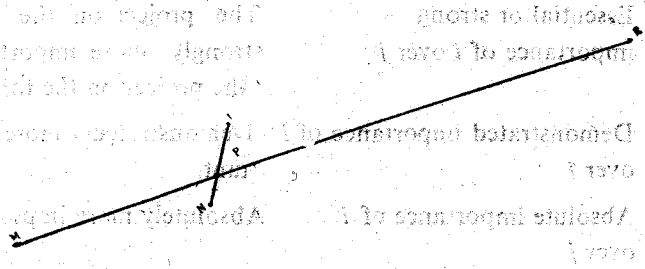

Figure 2
Relative Distances to Specific Points : Case 2

***K, L, M,* and *N* are four cities whose relative distances from city P are to be estimated.**

Case 2 : *Special Proximities*

The subjects were given the graphic in Figure 2 and similar judgement information as provided in Case 1.

Case 3 : *Relative Importance of Research Projects*

Finally, the subjects were given the following information :

Consider the following four research projects whose relative importances to soeiety are to be evaluated :

P_1—Research to find a cure for leukemia.

P_2—New missile development for national security purposes.

P_3—Research on the effects of birth control methcds on the user's health.

P_4—Research to develop fuel efficient internal combustion engines.

Record year judgments using the following scale :

Judgment Value	Definition	Explanation
1	Equal importance	The projects are of *equal importance.*
3	Slight importance of *i* over *j*	The project on the left is slightly more important than the project on the top.
5	Essential or strong importance of *i* over *j*	The project on the left is strongly more important than the project on the top.
7	Demonstrated importance of *i* over *j*	Demonstraiedly more important.
9	Absolute importance of *i* over *j*	Absolutely more important.
2, 4, 6, 8	Intermediate values between t'e two adjacent intensities.	

Reciprocals. If the project on the left is *less important* than the project on the top, use the reciprocals of the scale value. For example, if the project *j* (on the top) is slightly more important than the project *i* (on the left), one should enter 1/3 in cell (*i*, *j*).

Experimental Protocol

Prior to the experiment, the students were familiarized with the AHP Model. At the outset, the participants were given the GEFT.

After the completion of the GEFT, the participants were given the judgement tasks and asked to consider the three cases in the order presented. They were also asked to start the next case only after having completed the preceeding case, not to go back to the preceeding case(*s*), and not to use any measurement device. Even though the participants were familiar with AHP, in the beginning an example was given to make sure everyone understood the task. The subjects were told to ask questions if they were experiencing any difficulty. No subject indicated difficulty in completing the tasks.

4. RESEARCH HYPOTHESES

The first two cases were organized to test the judgement process for a special task. Given this, the first hypothesis is :

H_1 : The inconsistency attained for the first task will be lower than the inconsistency of the second task.

Hypothesis one is offered to "validate" the instruments since the second task is a more difficult task compared to the first task. While the size of the differential is not specified, the direction expected relates to the relative difficulty involved in rescaling for the second task. In the first case, the line segment ratio differentials directly yield values which are contained in the scale system. Therefore, if the subject estimated the differentials correctly, their scale values are immediately available. In the second case, the largest line segment is so large, that relative to the other line segments, the scale is not large enough to capture the proportional relationships if even relatively small numbers, such as 3, are assigned to line segments *PM*, *PL*, or *PN*. In this case, the subject must first estimate the ratios and then rescale them to orient them to the scale.

H_2 : Individuals who do well on the GEFT will render more consistent judgements than those individuals who do not do well on the GEFT.

In this case, the eigenvalue is a measure of the subject's ability to judge relative distances for the six pair-wise judgements. Since the

GEFT is a test of abstraction and orientation of special relationships, individuals who are skilled at such tasks are expected to be better at judging linear special differentials if that facet of cognition underlies the judgement rendering task.

No formal hypotheses are offered using the third case or the transference of transitivity across the tasks because directionality cannot be reasoned from prior research. For these cases exploratory analysis using the GEFT and demographic variables as well as examining the consistency of judgements over all three cases will be conducted.

RESULTS

The SAT and GEFT scores for the subjects are :

		Males ($n=32$)		*Females* ($n=33$)	
		Verbal	*Math*	*Varbal*	*Math*
SAT	Mean	43.7	47.9	46.6	46.6
	S.D.	7.4	10.1	7.0	10.8
GEFT	Mean	11.1		12.4	
	S.D.	5.2		4.4	

The inconsistency scores for the tasks are presented in Table 1.

TABLE 1

Inconsistency Scores

Inconsistency	Task 1	Task 2	Task 3
Mean	4.3	4.6	4.6
S.D.	.35	.51	.62
N	65	65	65

Using a matched pairs t-test the inconsistency scores for Task 1 are significantly different than the inconsistency scores for Task 2 $p<.0005$. Therefore, H_1 is supported by the data. It was more difficult for the individuals to render consistent judgements for Task 2 compared to Task 1.

Regarding H_2, the mean inconsistency scores for the high and low analytics are :

High* (n=32)				Low (n=33)			
Task	1	2	3	Task	1	2	3
$\bar{x}$	4.2	4.6	4.5	$\bar{x}$	4.2	4.6	4.6
S.D.	.46	.50	.60	S.D.	.18	.45	.64

* Individuals scoring greater than the median were classified as high analytics, others were classified as low analytics.

The inconsistency scores for the high and low analytics did not differ significantly $p<.1$. Further, the correlation coefficients for the inconsistency scores and the GEFT scores were not significantly different from zero for the three tasks. Therefore, H_2 is not supported by the data. A person's disembedding skill is not related to their ability to render consistent judgements.

The correlations of the inconsistency scores for the tasks are presented in Table 3.

TABLE 3

Correlations Across Tasks

		Task		
		1	2	3
Task	1	1	.14	.03
	2	—	1	.09
	3	—	—	1

The correlations in Table 2 are not significantly different from zero, suggesting that the level of inconsistency does not vary in a systematic way across tasks. (The second order correlations were also not significantly different from zero.)

Finally, sex and SAT scores were not factors in accounting for the differences among inconsistency scores.

5. DISCUSSION

These results suggest that there is no reason to expect that sex, disembeding skill or measures of verbal or mathematical scholastic apptitude are useful in predicting the ability to render consistent judgements using the analytic hierarchy. Further, the performance across tasks does not vary systematically. Therefore, in organizing groups or individuals to render judgements using the AHP, sex, disembedding skill, scholastic apptitude measured by the SAT, or previous performance rendering judgements using the AHP should not be used as screening criteria.

The fact that performance on a task does not suggest subsequent performance is an important result, because in our experience with the AHP people have a tendency to view performance in a particular task as suggesting that person's ability for subsequent tasks. If these people are excluded from subsequent analyses in an attempt to develop a cohort of experts in using the AHP, the group of individuals rendering judgements will likely become too small to attain the reasonable diversity of perspectives required to stimulate a comprehensive dialogue.

REFERENCES

1. M L. Bariff and E.J. Lusk, Cognitive and Personality Tests for the Design of Management Information Systems, *Management Science*, April 1977, pp. 820-829.

2. I. Benbasat and A. Dexter, Value and Events Approaches to Accounting : An Experimental Evaluation, *Accounting Review*, October 1979.

3. R.H. Doktor and W.T. Hamilton, Cognitive Style and the Acceptance of Management Science Recommendations, *Management Science*, April 1978, pp. 884-894.

4. H.F. Hess, Additional Evidence on the Generality of Cognitive Complexity, *Perceptual and Motor Skills*, Vol. 23 (1966).

5. J.H.B.M. Huysmans, The Effectiveness of Cognitive Style Constraint in Implementing Operation Research Proposals, *Management Science*, September 1970, pp. 92–104.

6. E.J. Lusk, Investigation of the Surrogation of Judgemental Inconsistency with (Lamda-N)/(N-1), Working paper, Wharton School, University of Pennsylvania, 1979.

7. T.L. Saaty, A Scaling Method for Priorities in Hierarchical Structures, *Journal of Mathematical Psychology*, Vol. 15 (1977).

8. T.L. Saaty, The Analytic Hierarchy Process, McGraw-Hill, Inc., New York, NY, 1980.

9. L.G. Vargas, A Note on the Eigenvalue Consistency Index, *Applied Mathematics and Computation*, 1980.

10. H. Witkin, R. Dyk, H. Faterson, D. Goodenough and S. Karp, Psychological Differentiation, Wiley, New York, 1962.

11. D. Goodenough Witkin and S. Karp, Stability of Cognitive Style from Childhood to Young Adulthood, *Journal of Personality and Social Psychology*, Part I, Vol. 7, November 1967.

12. P.K. Oltman Witkin, E. Ruskink and S.A, Karp, The Embedded Figures Test, *Consulting Psychologist Press*, Palo Alto, CA, 1971.