

FEWER COMPARISONS – EFFICIENCY VIA SUFFICIENT REDUNDANCY¹

William C. Wedley

Simon Fraser University & W. C. Wedley and Associates Ltd.

8888 University Drive

Burnaby, B. C., CANADA, V5A 1S6

wedley@sfu.ca

ABSTRACT

“Too many comparisons” is a frequent complaint expressed by AHP/ANP users. This paper proposes a methodology to reduce the number comparisons while still allowing complete redundancy. It involves 4 steps: (1) Roughly rank the objects; (2) Using the lowest ranked object as the unit, estimate intensities for each of the other objects; (3) Using other objects as units, make sufficient additional comparisons until predicted consistency is acceptable, (4) Allow the user to continue with more comparisons if so desired. At the end of stage 2 or thereafter, the user can terminate comparisons. Besides allowing fewer comparisons, ranking before comparing holds out promise to reduce coding errors and improve accuracy.

Keywords: Pairwise comparisons, redundancy, predicted consistency, efficiency

1. Introduction

Analyzing a problem with AHP/ANP generally involves three steps: (1) Structure the problem, (2) Compare the components and (3) Synthesize the results. The first stage is perhaps the most important. By decomposing the problem into clusters of alternatives, criteria, scenarios, or other perspectives, we get to understand the major factors affecting the problem. The second stage is the grunt work. Laboriously, we make multiple sets of comparisons within and between clusters so as to derive ratio scales of impact, likelihood, or preference. The final stage is generally programmed according to our choice of aggregation.

This paper is concerned with the second stage – particularly the process of making one set of paired comparisons. Although efficiency (fewer comparisons) is the key focus of this study, ease of use, accuracy, and the nature of the problem are other important attributes (Millet, 1997). The main objective herein is to structure the process into stages that allow the decision maker (DM) to make fewer comparisons before deciding whether priorities are acceptable.

We use the AHP/ANP concepts of decomposition to break the comparison process into stages. Each stage can be analyzed and scrutinized for improvements. The recommended process involves four stages:

- (1) Roughly rank the items.
- (2) Select a referent object as the unit of measure. Establish intensities for the other n-1 objects
- (3) Make sufficient additional comparisons until predicted consistency is within an acceptable range
- (4) Allow the user to continue with more comparisons if so desired.

After stage 2, priorities can be displayed. The DM uses his/her judgment to (1) accept the displayed priorities or (2) continue making comparisons. Ultimately, it is the DM's choice to continue or not.

The next section is devoted to a description of the problem. Section 3 is devoted to a discussion of related prior literature. Section 4 goes into the details of the 4 stages. It illustrates how each stage could be activated within a computer program. Section 5 addresses some common concerns that may be expressed

¹ The Natural Sciences and Engineering Research Council of Canada provided the financial support necessary for this project.

with the procedure. Section 6 presents some evidence of how well the procedure would work in practice. It compares the results at various stages with the true known values. Section 7 discusses advantages and disadvantages of the proposed technique and Section 8 provides some concluding remarks.

2. The Problem – Too many comparisons

A common complaint with AHP/ANP is that the comparisons become tedious. Users say that too many comparisons are required, that they soon get tired of the repetition and that they lose attention. Others complain about the time involved. Although this problem was discussed in a plenary session of the 2005 International Symposium on the Analytic Hierarchy Process, little progress has been made. Instead, adaptive mechanisms are used. This writer knows one AHP/ANP expert who prefers direct estimation of n values. He then tinkers with those values until he is satisfied with the priorities.

Consider an AHP problem with seven alternatives and six criteria. For each criterion, $n(n-1)/2 = 21$ comparisons are required, each asking the following two questions:

- (1) Of the two alternatives being considered, which is more dominant? (This ordinal question identifies the less dominant object that will be used as the unit of measure for the comparison).
- (2) By how many times or how much more dominant is that alternative? (This is an intensity question that determines the absolute value of the ratio between the objects).

Since two questions are required for each comparison, 42 questions are required for each criterion and another 30 for the criteria. For the overall hierarchy, $42 * 6 + 30 = 282$ questions are required. If the problem is being evaluated by a group of 10 individuals, the total number of questions increases to 2820.

The time to perform those comparisons can become excessive. In a study using AHP to estimate the distances of 8 cities from Singapore, this researcher discovered that the average time for subjects to complete 28 comparisons was 17.3 minutes (37 seconds per comparison). Millet (1997) found that the average time for students to compare 5 geometric objects was 5.9 minutes (35 seconds per comparison). If 36 seconds are required for each comparison in our sample hierarchy, each individual would require 85 minutes to complete the task. If 10 individuals participate, the total time rises to 14.1 hours.

3. Related Literature

Literature related to improving the comparison process falls into four categories: incomplete comparisons, stopping rules, starting rules and elicitation methods.

3.1. Calculating priorities with incomplete comparisons

Harker (1987a, b c), an early advocate of reducing the number of comparisons, has provided the most comprehensive suggestions. He developed two methods for calculating missing values of the incomplete comparison matrix. The first is based upon the geometric mean of all paths to the missing value (1987c). Since the connecting paths grow long and the computations difficult, this method has been judged to be less practical (Setiawan, 2002; Fedrizzi & Giove, 2007). Harker's second method (1987b) replaces missing values by w_i/w_j , the unknown vector values to be estimated. He demonstrated that this formulation could be solved as the eigenvector of an augmented matrix that has the number of row missing values plus one in the diagonal. At about the same time, Takeda and Yu (1988) came to the same conclusion.

As well as providing calculation routines for incomplete comparisons, Harker (1987a) devised a gradient method for choosing the next comparison that would add the most information and stopping rules for terminating the comparison process. He proposed generating n comparisons with $n-1$ of them forming an

interconnected spanning tree. Priorities are calculated and the gradient procedure is used to pick the next comparison. This process repeats until the DM decides to stop. Harker suggested that stopping could be based upon the DM's satisfaction with the weights, a tolerance change in weights between steps, or predicted ordinal stability in the weights.

Shen et al (1992) took a different approach to reducing the number of comparisons. They estimated the vector by breaking it down into subsets and linking via a common object. In this manner, complete comparisons are applied to smaller matrices that are then linked by the common object. This approach is similar to Saaty's concepts of linking between orders of magnitude or across many alternatives.

To test the effect of incomplete comparisons on accuracy, Carmone et al (1997) simulated large sized matrices (10, 15, and 20) and then used various methods of deletion until they got to $n-1$ comparisons. They discovered that about 57% of the redundant comparisons could be removed without causing the priorities to be significantly different from the full matrix results. They also discovered that even greater reduction is tolerable if large magnitude comparisons are maintained and low magnitude ones discarded. Although their matrix sizes were outside what is normally used in AHP, their results provide support for the proposition that not all comparisons are necessary.

3.2. Starting Rules

Before calculating priorities from incomplete comparisons, it is necessary to generate at least $n-1$ comparisons that form an interconnected spanning tree. Wedley et al (1993) investigated starting rules for selection of the first $n-1$ comparisons. They conducted an experiment with 144 business students for a problem with a known answer. The students' task was to estimate proportions of five distinct colors in a rectangular area under different experimental conditions. Six different referents were considered for the first $n-1$ comparisons: vertical (first column), horizontal (bottom row), upper diagonal, lowest ranked item, median item and highest ranked item. The lowest ranked item for the first $n-1$ comparisons proved to be statistically more accurate than any of the other starting methods. As well, the results from the lowest priority item were more accurate than the priorities from complete comparisons.

Other people used different methods to select the starting comparisons. Harker (1987b) used random selection. Ishizaka and Lusti (2004) devised an expert system that assists the DM to derive a comparison matrix that is within a tolerable level of inconsistency. Their starting rule was to select the comparisons which formed the upper diagonal of the comparison matrix. Ra (1999) also selected $n-1$ diagonal comparisons plus one extra comparison from the upper corner of the matrix. This enabled him to do an immediate consistency check for his chainwise process.

Setiawan (2002) in a Master's thesis studied five different methods: one alternative (row or column), diagonal comparisons after ranking alternatives, random, highest ratio comparisons, and median ratio comparisons. Like Carmone et al (1997), Setiawan used large simulated matrices (10, 15, and 20). He discovered that ranking followed by diagonal selection was the best method and that just $n-1$ comparisons provided at least 97% accuracy. Being simulated, not behavioral, these results are subject to the vagaries or randomness. The high accuracy is associated with the large n and minimal perturbing of errors.

3.3. Stopping Rules

As noted above, Harker (1987a) suggested three different methods for stopping the incomplete comparison approach. Wedley (1993), working with Harker's incomplete and gradient approaches, suggested another method based upon an incomplete consistency ratio. Matrices with different degrees of inconsistency were simulated and these were then used to develop regression equations that predict the consistency ratio at each step beyond n comparisons. This information, plus the upper level of the 95 confidence interval, is then incorporated into series of decision rules that help determine whether the DM

can stop making comparisons. Essentially, if the predicted consistency ratio is too high (>0.2), the DM is required to continue making comparisons. But if both the predicted consistency ratio and confidence interval are small, the priorities are displayed and the DM is given the option to discontinue comparisons.

3.4. Elicitation formats

Research on how comparisons should be elicited is sparse. Webber et al (1996) used questionnaire input to test scales (numerical, verbal, graphical), formats (random vs. non-random) and order (top down or bottom up). They discovered little evidence of scale or order effect and some evidence of format effect. Unfortunately, their approach was incapable of showing which effects are better.

The most comprehensive study on elicitation was carried out by Millet (1997). His experimental design involved estimating the area of five objects commonly used to validate AHP (shown in Figure 1 below). He looked at 5 different elicitation modes available with Expert Choice: (1) direct estimation (2) graphical multiple bars (3) graphical pairwise bars, (4) numeric pairwise comparison, and (5) verbal pairwise comparison. Graphical moving bars proved to be easier to use but they were not as accurate. Numeric and verbal comparisons produced the most accurate results, but they were perceived as difficult methods that took more time to perform. Because it appeared that relative accuracy was associated with cognitive orientation (verbal vs. quantitative), Millet suggested that maybe the elicitation method should be contingent upon factors such as the orientation type. He did recognize that there is a tradeoff between greater accuracy (verbal and numeric modes) and easier use (graphic bars).

Because there is ample evidence that people have very different interpretations of verbal expressions, Huizingh and Vrolijk (1997) decided to compare the verbal and numerical modes of assessing judgments. They concluded that the numerical mode is more accurate and with greater consistency, but the difference is not significant. They suggested that it may be wise to help users by showing the conversion table that is used to change verbal expressions to numeric values.

4. The Proposed Elicitation Method

The proposed method combines ideas from the previous literature. The starting rules of Wedley et al (1993) are used to first rank the objects and then select $n-1$ comparisons (Steps 1 & 2). Harker's gradient method (1987a) is then used to pick subsequent comparisons and his incomplete method (1987b) is used to calculate priorities (Steps 3 & 4). Wedley's (1993) predicted consistency is used as a device to determine sufficient redundancy levels and when the DM can potentially stop (Step 3). Adjustable bars are used because Millet (1997) found those to be the easiest to use. Also adopted is the suggestion of Huizingh and Vrolijk's (1997) to display both the numeric and verbal values alongside the fundamental scale. This avoids misunderstandings and maintains potential accuracy (Millet, 1997; Huizingh & Vrolijk, 1997). The following steps demonstrate the melding of concepts.

4.1 Step 1: Rank the objects to be compared

Before actually undertaking the paired comparisons, the first step should be to rank the objects that are being considered. A pictorial analogy for this step is presented in Figure 1. The left side of the figure presents five shapes commonly used to validate AHP. On the right, the same shapes have been ranked in decreasing order of magnitude. Since ranking is an easier cognitive task than estimating magnitudes, this preliminary step helps structure the comparison task in the DM's mind. In subsequent steps the DM can concentrate on the magnitudes without worrying about potential errors like ordinal inconsistencies (e.g. inadvertently coding the rectangle is as being bigger than the diamond). As well, there will be fewer ordinal questions. Whereas regular comparisons require $n(n-1)/2$ ordinal questions, prior overall ranking requires only n questions.

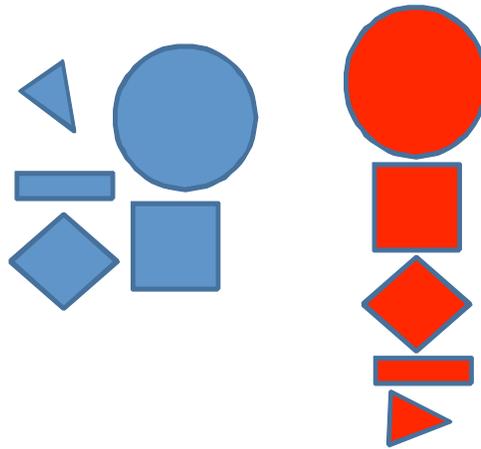


Figure 1. Random Order vs. Rankings of Shapes

In most AHP problems, we do not have physical objects to compare. Instead, the objects are mental constructs like strategies, likelihoods or preferences. This fact, however, does not prevent us from making ordinal questions beforehand. We could ask questions such as

1. Which item is the least dominant
2. Which item is most dominant
3. Which item is the next most dominant, etc.

In a computer program, Figure 2 illustrates how this process could take place for the unranked objects of Figure 1.

STEP 1. Roughly rank the alternatives (best at top, worst at bottom)

The objective is to establish relative priorities that sum to one. Initially, they are zero.

First, we rank the alternatives. To the right are the unranked alternatives. Please answer the questions by entering the unranked alternative number in the Answer column. If two alternatives are tied, it does not matter which order they are entered.

Priorities	Ranked	Questions	Answers	Unranked
0.000	1. Circle	Which is best (strongest)?	1	1. Circle
0.000	2. Square	Which is 2nd best?	3	2. Triangle
0.000	3.	Which is 3rd best?		3. Square
0.000	4.	Which is 4th best?		4. Diamond
0.000	5. Triangle	Which is worst (weakest)?	2	5. Rectangle
1.000				

Figure 2. Computer output of the ranking process of Step 1.

Once the ranking is completed, the items can be displayed as equal priorities. They would be equal because we have not yet rated them for intensity. Figure 3 illustrates the ranked output at the end of Step 1. It shows the fundamental scale that will be used in subsequent stages to evaluate intensities.

With the objects ranked, we can proceed to establish intensities for the relationships. This stage involves selection of a starting rule for the first n-1 comparisons. As Ishizaka and Lusti (2004) point out, only the first n-1 comparisons (if forming a spanning tree) are truly independent of transitivity relationships. Our objective is to select n-1 comparisons that are intuitively logical, easy to compute, and accurate.

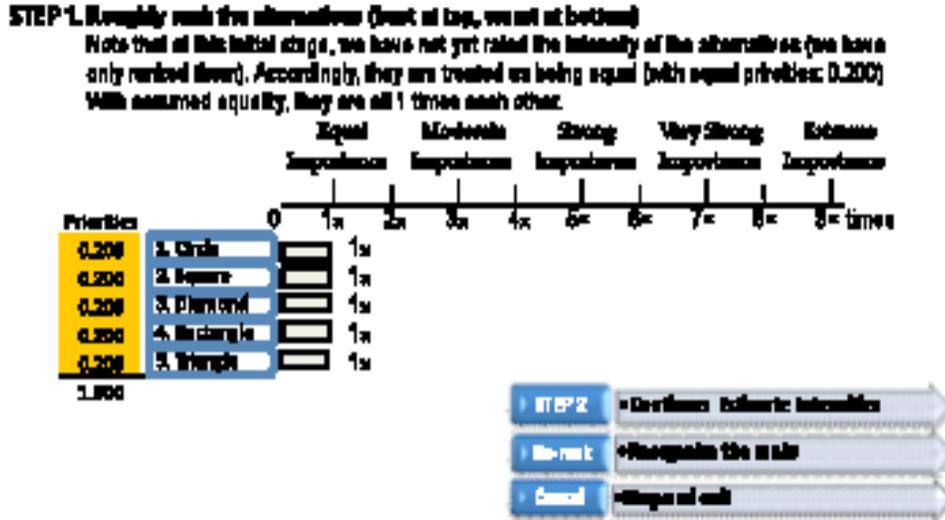


Figure 3. Computer output of ranked items at the end of Step 1.

4.2 Step 2: Establish Intensities for the first n-1 comparisons

Whitaker (2007) has reported how one individual evaluated the non-ranked shapes in the left of Figure 1. It is useful to refer to this person’s comparison matrix when explaining how subsequent comparisons are chosen. Notice in Figure 4 that the individual did not rank the objects beforehand.

	Circle	Triangle	Square	Diamond	Rectangle	Eigenvector Priorities	Actual True Priorities
Circle	1	9	2.5	3	6	0.488	0.471
Triangle	1/9	1	1/5	1/3.5	1/1.5	0.049	0.050
Square	1/2.5	5	1	1.7	3	0.233	0.234
Diamond	1/3	3.5	1/1.7	1	1.5	0.148	0.149
Rectangle	1/6	1.5	1/3	1/1.5	1	0.082	0.096

Figure 4. One Individual’s Comparison Matrix for areas of 5 objects (Whitaker, 2007)

The recommended starting rule for the first n-1 intensity estimates is to make the first n-1 intensity estimates with the lowest ranked object as the standard referent. Several reasons exist for the choice of this unit of measure. Of the non-reciprocal comparisons, the lowest ranked object is the only one that is the common unit for comparisons with all other objects. In Figure 4, the unit of measure for each column is listed at the top of the column. Notice that Triangle is the only object that does not have a derived reciprocal value. Because all other objects have greater intensity than the lowest ranked object, the cognitive benefits of ranking can be maintained. Any other n-1 spanning tree (e.g. first row or column, upper diagonal) will require a change in thought process as the dominated object changes.

A second reason for choosing the lowest ranked object is accuracy. As noted in the literature review, Wedley et al (1993) discovered that the lowest ranked item for the first n-1 comparisons would produce statistically more accurate results than other starting rules. They put forward the idea that it may be easier to compare when the least important item is in the denominator rather than vice versa. As well, they

speculated that the lowest priority item is easier to compare because it provides the greatest difference in magnitude. Using simulation, Carmone et al (1997) came to a similar conclusion that larger magnitude comparisons yield more accurate results. Although this evidence is limited and requires replication, it provides some support for using the lowest ranked object for the first n-1 intensity estimates.

Several authors have considered how to select the initial comparisons. Harker (1987a) used a random approach, whereas Ra (1999) and Ishizaka & Lusti (2004) opted for a spanning tree of diagonal items. Ishizaka and Lusti (2004) rejected a common row or column approach because they felt that it compromised the psychological independence of the comparisons (which they considered to be an advantage of AHP). Although comparison to the lowest ranked object is a common column approach, it is an open question as to whether psychological independence between comparisons is a desirable attribute that outweighs structure that comes from ranking the items.

The suggested method for estimating intensities is presented in Figure 5. Initially, the two objects to be compared are presented in a common color (yellow) that is different from the other objects. As soon as the bar of the more dominant common is grasped and dragged, the color of the less dominant item changes to red to indicate that it has been set as the unit of measure. For the first n-1 comparisons of Step 2, Triangle should be the fixed unit, because it is the lowest ranked. In Figure 5, Square is the last object to be evaluated during Step 2. Its magnitude is judged to be 5 times larger than the Triangle. Notice that the previous comparisons to Triangle are shown in a muted color. This display of other comparisons is only maintained for the first n-1 comparisons.

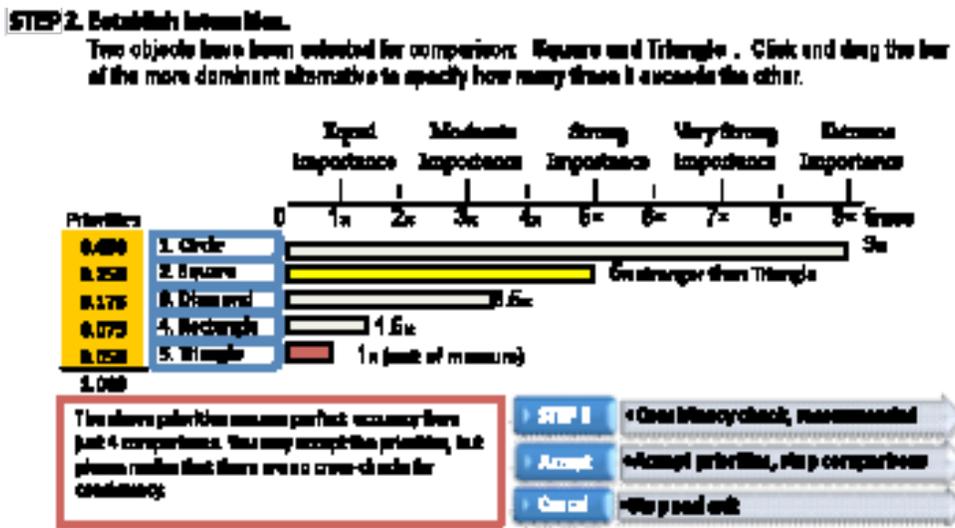


Figure 5. Computer output at end of Step 2.

During this second step, the values automatically adjust in the Priorities column whenever a bar is moved. At the end of Step 2, the DM has the choice of accepting the priorities, continuing to more comparisons or exiting the process. At n-1 comparisons, the DM is advised that cross checks are impossible and that the comparison process should continue.

4.3 Step 3: Consistency Check

Beyond n-1, the comparisons are no longer independent of each other. Through transitivity relationships, we can impute the value of missing comparisons and calculate the consistency between comparisons. If the divergence between the input and imputed value is large, then the DM can be advised.

When to allow the DM to stop making comparisons is more complex. Because it is ultimately the DM's choice as to which priorities are acceptable, the method proposed herein is to use DM discretion bolstered by Wedley's (1993) decision rules associated with predicted consistency and its 95% upper confidence interval. In effect, stopping comparisons is based upon the DM satisfaction with the priorities bolstered by the fact that they have been generated by a relatively consistent process.

Figure 6 shows the results that would be presented to the DM after sufficient redundancy. The last comparison the DM performed was the sixth one chosen by Harker's gradient method (Diamond vs. Rectangle). Notice that the DM is invited to progress to Step 4. This occurs only because the DM's predicted consistency ratio is low (0.027), with a 95% probability of being less than 0.079.

It should be re-iterated that the comparison display is not the same as in Step 2. Because the referent alternative keeps changing (it is a function of Harker's gradient method), it is not possible to show previous comparisons in terms of a standard unit. Therefore, all non-compared alternatives are set at the neutral unit value. If the DM does decide to accept the priorities and stop comparisons, then the computer output will give a final display similar to Figure 5 (bars adjusted so least ranked is the unit) or Figure 7 below (scale changed and bars adjusted to priority values).

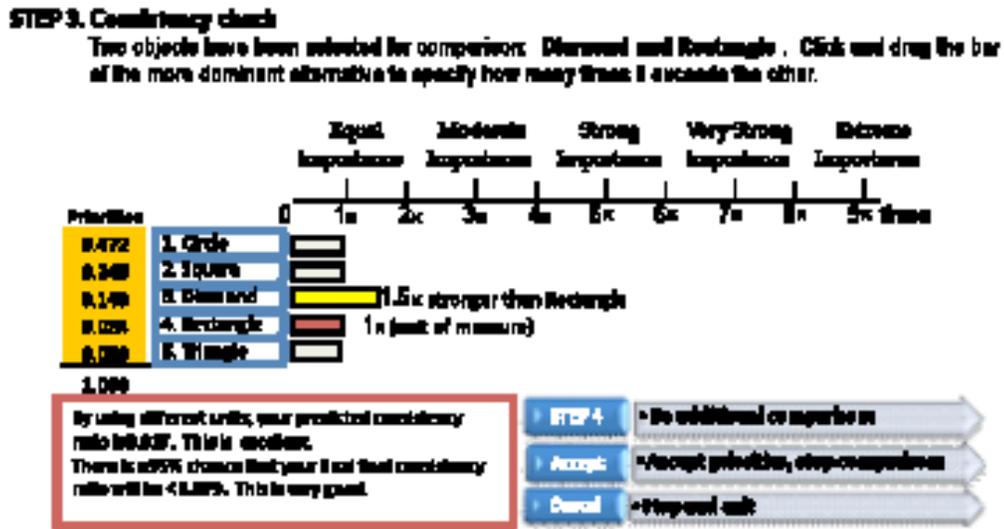


Figure 6. Computer output at end of Step 3.

4.4 Step 4: More comparisons

Figure 6 shown above will never be displayed if the predicted consistency ratio is high (>0.2). Instead, the computer will automatically request additional comparisons. Because the predicted consistency ratio is low when Figure 6 is finally displayed, it is likely that the DM will accept the results and stop comparing.

The choice, however, is up to the DM. He/she may continue with Step 4 and do more comparisons. The display screen will remain the same as Figure 6, except that the priorities and predicted consistency figures will change. Provided the DM does not make a radically inconsistent comparison, the predicted consistency ratio will change modestly and its confidence interval will get smaller. If all comparisons are completed, the eigenvector priorities will be the same as shown in Figure 4. Figure 7 shows those final priorities after complete comparisons. An alternate display is to show the priorities with the lowest ranked object as the unit. That display would be similar to Figure 5.

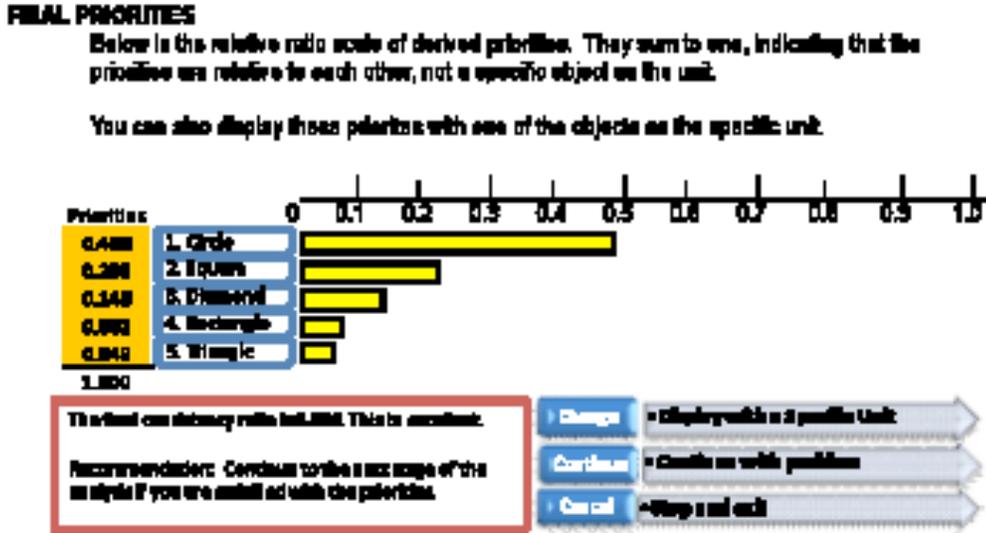


Figure 7. Computer output at end of all comparisons

Figure 8 gives a comparison of the accuracy of priorities after n-1 comparisons, sufficient redundancy (6 comparisons) and complete (10) comparisons. It is apparent that this DM should have done more than n-1 comparisons, but not all comparisons.

Step	Circle	Square	Diamond	Rectangle	Triangle	Mean	Mean
						Absolute Error	Percent Error
After 4 Comparisons (Step 2)	0.450	0.250	0.175	0.075	0.050	0.017	10.1%
After 6 comparisons (Step 3)	0.472	0.245	0.149	0.084	0.050	0.005	3.5%
After all 10 Comparisons	0.488	0.233	0.148	0.082	0.049	0.007	4.3%
True Priorities	0.471	0.234	0.149	0.096	0.050		

Figure 8. Comparison of accuracy at different number of comparisons.

5. Issues and Concerns with the Proposed Elicitation Method

5.1 Potential Errors in the Ranking Process

Step 1 is the “roughly rank” stage. What happens if the DM ranks too roughly and the ranks are incorrect? For example, Triangle and Rectangle have similar sizes – Triangle may be ranked 4 ahead of Rectangle.

This type of ordinal error should be picked up in the subsequent stages when intensities are assessed. For instance, during Step 2, the DM would be asked to compare 4.Triangle to 5.Rectangle. Sensing Rectangle as being dominant, the DM is likely to adjust that bar, leaving the Triangle as the unit of measure for the comparison. When this is done, the Triangle intensity bar 4 turns red (signifying it is the unit). During step 2, such a red unit bar above the adjusted bar signifies a conflict with what was done in step 1. A warning message can then be issued asking the DM whether he/she wants to change the original ranking.

5.2 Ranking may preclude legitimate reversals

A second concern is that prior ranking eliminates opportunities for reversals, particularly when some objects are copies or near copies. While it is true that prior ranking will make reversals less likely, it is not true that the process will eliminate reversals.

At Steps 3 and 4, whenever a red unit bar is above the adjusted bar, it signifies that a potential ordinal error has been made. In such circumstances, the following type of warning message can be given:

You previously said Triangle > Rectangle. Now you say Rectangle > Triangle. Which is it? Do you want to switch the order?

If the DM decides that the order should be reversed, then that will happen.

Reversals in the order of alternatives can also occur as the result of inconsistency in transitive relationships. Again, such reversals would be allowed if the DM agrees that it should happen. The recommended procedure to do this is to give the DM a reminder notice whenever the latest re-calculation of priorities signifies that the ranks have changed. The DM would then be given to power to allow the change or to keep the ranks in their current order.

5.3 Stopping after n-1 comparisons is risky

Preliminary ranking reduces the probability of an ordinal error at the outset of the comparison process. Accordingly, priorities based upon n-1 should be quite accurate. Making redundant comparisons (> n-1) provides cross checks for consistency, but there is no guarantee that priorities based upon >n-1 are more accurate than n-1. The ultimate decision maker of the appropriateness of the priorities is the DM. If he/she decides to accept priorities after just n-1 comparisons, then the analyst should respect that decision.

At the end of Step 3 and thereafter, the priorities are displayed for the DM after each new comparison. It is possible that the DM may prefer the priorities of a previous step. It would be relatively easy to give the DM this option. We could give the DM all sets of priorities from n-1 to the current comparison and let the DM decide which set is best. Again, the ultimate decision maker is the DM.

6. Tests of the Proposed Method

To test the proposed method, it is desirable to use data with a known result. Such data are available for 42 university students who estimated the distance of cities from Singapore. This experimental condition is similar to the distance from Philadelphia problem that was used to illustrate AHP (Saaty, 1977).

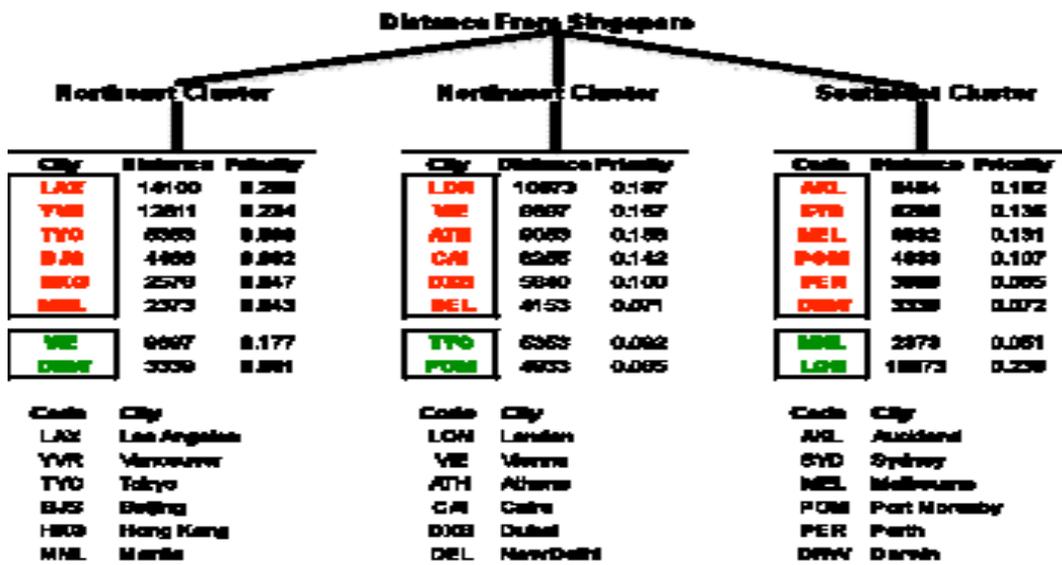


Figure 9. Template 1 of 6 for Estimates of Distances from Singapore

A total of 18 cities were chosen for analysis. Six of them were in a northeast direction from Singapore, 6 from a northwest direction and six in a southeast direction. Since the objective was to derive a single overall scale of distance from Singapore, a hierarchy was constructed with the cities of each direction as a cluster. One city from each of the other two directions was included in each cluster so that linkages could be established between clusters. Figure 9 presents one of six templates that were evaluated.

When making evaluations, each student had a globe of the world which had the cities tagged. The student could look at the globe and even spin it, but he/she could not use any device such as a string to measure between cities. Each student had to evaluate 3 clusters composed of 8 cities. Since each cluster required 28 comparisons, each person made 84 comparisons. A money prize was given to the person with the most accuracy. The only differences between templates were the cities used as links between clusters.

Figure 10 illustrates the comparison matrix of one student for the northwest cluster of Figure 9. Note that except for the linking cities (TYO - Tokyo and POM - Port Moresby) the cities are ranked from the farthest away to the closest. In spite of such ranking, the computer program that was used always took the first n-1 comparisons in reference to the most distant of the ranked cities (i.e. row 1). Accordingly, the data set does not enable direct evaluation of Steps 1 and 2.

Northwest Cluster of Template 1

Distance	City	MAX		MED			MIN	NE	SE	Eigenvector Priorities	True Priorities
		LON	VIE	ATH	CAI	DXB	DEL	TYO	POM		
10873	LON	1	1.1	1.4	1.6	1.9	2.3	2.1	2.2	0.194	0.187
9697	VIE	0.91	1	1.1	1.1	1.5	2.2	1.9	1.9	0.164	0.167
9053	ATH	0.71	0.91	1	1.1	1.4	2.1	1.7	1.8	0.150	0.156
8255	CAI	0.63	0.91	0.91	1	1.3	2.1	1.4	1.4	0.135	0.142
5840	DXB	0.53	0.67	0.71	0.77	1	1.6	1.2	1.1	0.106	0.100
4153	DEL	0.43	0.45	0.48	0.48	0.63	1	0.91	0.91	0.074	0.071
5353	TYO	0.48	0.53	0.59	0.71	0.83	1.1	1	1.1	0.090	0.092
4933	POM	0.45	0.53	0.56	0.71	0.91	1.1	0.91	1	0.087	0.085

Figure 10. One individual’s comparisons for the Northwest cluster of Template 1

After the students reached sufficient redundancy conditions of Step 3, the computer recorded their priorities, the predicted consistency index and the 95% upper confidence interval at that stage (usually after 11 paired comparisons). Accordingly, we can determine their accuracy at that stage and at the end of all 28 comparisons. As well, we can look at referents for the first n-1 comparisons so as to simulate the results of different referents for starting rules of Step 2. Five referents were selected:

1. Maximum (farthest) city as the common referent (MAX in Figure 10)
2. Minimum (closest) city as the common referent (MIN in Figure 10)
3. Median city as the common referent (MED in Figure 10)
4. Linking City 7 as the common referent (NE – North East in Figure 10)
5. Linking City 8 as the common referent (SE – South East in Figure 10)

To test the efficacy of different steps, all estimations can be compared to the true priority values. Hence we have the following error measures:

1. Five different referents for the first n-1 comparisons (equivalent to different Step 2 outcomes)
2. Sufficient redundancy (equivalent to the Step 3 result)
3. Full redundancy (equivalent to all comparisons at the end of Step 4).

Mean Absolute Error (MAE) and Mean Absolute Percent Error (MAPE) are used to measure errors from the true priorities.

Table 1 presents the error measures for the various methods. Notice that both the MAE and MAPE values are highest when just n-1 comparisons are performed. Using the closest and median city as the referent produced the lowest degrees of error amongst the n-1 methods. At sufficient redundancy (11 comparisons), the error measures are lowered. However, they are not as low as the results with complete comparisons (28). The general conclusion is that both MAE and MAPE improve with more comparisons.

For at least two individuals, that finding does not hold. At complete comparisons, they had a Consistency Ratio >0.1 whereas their predicted CR at sufficient comparisons was much lower. Upon closer inspection, it was revealed that those individuals committed ordinal errors during Step 4 by saying one city was farther away when it was not. In other words, their inaccuracy increased with more comparisons although they did not realize it. Perhaps the ranking of Step 1 would have helped to avoid such errors.

Table 1 Summary of MAE and MAPE results

One Variable Summary	City used as Referent for n-1 Comparisons					Sufficient	Complete
	Closest	Farthest	Median	City 7	City 8		
Number of Comparisons	7	7	7	7	7	11	28
Mean Absolute Error	0.013	0.014	0.013	0.015	0.015	0.012	0.009
Std. Dev.	0.006	0.011	0.008	0.008	0.012	0.008	0.005
Mean Absolute Percent Error	11.3%	13.0%	11.0%	12.7%	13.1%	10.9%	8.1%
Std. Dev.	5.0%	10.0%	7.2%	7.0%	13.5%	8.0%	4.6%

Paired t-tests were conducted on alternate referents for the n-1 comparisons of Step 2. For MAE, no statistically significant differences were found between the two best referents, the closest and median cities. Both of those cities were significantly better than Linkage City 7 ($p < .05$) and the median city was modestly significantly better than the farthest city ($p < .10$). Similar results were found on the paired t-tests based upon MAPE. Both the closest and median cities were insignificantly different from each other, but significantly better than the farthest city and City 7 ($p < .05$). While closest city is no better as a referent than the median city, the insignificance between the two and their superiority over other methods tends to support the lowest ranked object for the unit of measure for the first n-1 comparisons.

After 11 of 28 comparisons, all subjects met the sufficient redundancy conditions of Step 3. To determine whether 4 extra comparisons beyond n-1 resulted in an improvement of accuracy, paired t-test were conducted on the MAE and MAPE values for sufficient comparisons vs. those of the closest and median cities as referents. Although the additional comparisons yielded more accurate priorities, the differences were statistically insignificant ($p > 0.15$).

Paired t-tests for differences in means were also conducted to determine whether fewer comparisons were worse than complete comparisons. For both MAE and MAPE, the differences from results with complete comparisons in Figure 1 are statistically better ($p < .001$) than any of the other differences. The implication is that going the extra distance to complete comparisons yields greater accuracy. Not determined is whether some point between 11 and 28 comparisons is better or insignificantly different from full redundancy. Only a reanalysis of the data will allow that.

We can determine, however, that extra comparisons come at a cost. The average time to complete the first 11 comparisons was 8.01 minutes or 44 seconds per comparison. The average time to complete the remaining 17 comparisons was 9.24 minutes (33 seconds per comparison). The longer time for the earlier comparisons is probably associated with the student becoming familiar with the task.

Even though the improvement between sufficient redundancy and complete comparisons is statistically significant, is it worth spending over nine minutes to improve the accuracy of the typical priority by 0.003 or 2.8%?

7. Discussion

In problem solving, the concepts of decomposition, structure, analysis and synthesis are typical tools of the trade. Indeed, those concepts embedded in AHP/ANP have proved to be very beneficial. But those concepts have not been adequately applied to AHP/ANP's perennial problem of too many comparisons.

This paper attempts to provide decomposition, structure and analysis to the paired comparison process. The first step, to roughly rank the objects, establishes structure to the items before magnitudes are estimated. Although ranks can be imprecise (they can be corrected at subsequent steps), the word "roughly" does not imply that the ranking should be done in a cavalier or haphazard manner. Rather, it implies that Stevens' (1946) assignment of numbers to objects according to order is a coarse type of information. In order to measure something in the sense of calibration, quantity and intensity, we need further knowledge of magnitudes and the degree one unit exceeds another (Michell, 1997).

Although ranking does not give us true, quantitative measurement, it can be a precursor and aid to achieving such measurement. Ranking is an easier cognitive task than determining intensities. If we rank our objects so that they line up in natural progression, then it is subsequently easier to determine the intensities between them. This is the rationale for Step 1: it captures the advantages of ranking, and eases the task of estimating intensities.

Some, like Ishizaka and Lusti (2004), may say that psychological independence should be enforced and that comparisons should be mixed up so they look dissimilar and random. The difficulty with that perspective is that it raises the cognitive load on the user and it ignores the fact that comparisons must somehow be interconnect in a spanning tree. As well, looking at just 2 items at a time never takes advantage of Miller's 7 ± 2 rule for the number of objects an individual can effectively keep in mind (Miller, 1956, Saaty & Ozdemir, 2003). By allowing all objects and intensities to be seen at one time, this psychological strength can be accentuated.

Prior evidence (Wedley et al, 1993) was the major reason for picking the least ranked object as unit for the first $n-1$ comparisons. In that earlier study, the least ranked color turned out to be the most accurate referent. In the distance from Singapore study reported herein, the lowest ranked city and the median city were about equal in ability as initial referents. The largest ranked city was not. While evidence is still limited, it provides modest support for using the least ranked item. We should note that in both the Singapore study and Wedley et al (1993), objects were not ranked beforehand. In any future studies to assess starting rules, prior ranking should be one of the experimental conditions.

Step 2 of the proposed elicitation method is much like Stevens' magnitude estimation (Stevens, 1971). The essence of Magnitude Estimation is to present the evaluator with a standard stimulus (a modulus) with a specified magnitude value such as 10. The evaluator then considers other objects and assigns numerals to them that are relative in intensity to the standard stimulus. As Narens (1996) observed, any object of the choice set could be the standard stimulus and could be given the number 1 as the subjective norm. That is much like Step 1 where the lowest ranked object is given a standard stimulus numeral equal to 1. Other objects take their intensities from that unit. From this perspective, magnitude estimation can be envisaged as establishing one column of the paired comparison matrix. If the DM is confident and satisfied with the priorities after $n-1$ comparisons, then stopping there is much like magnitude estimation.

At Steps 3 and 4, AHP/ANP goes beyond the simple scaling of magnitude estimation. AHP/ANP uses other objects as the unit of measure. Whereas magnitude estimation tends to operate under assumed error-free conditions, AHP/ANP embraces the concept of moderate error and the opportunity to measure it through redundant comparisons.

Although the DM is allowed to stop and accept priorities at the end of Step 2, the recommendation at that stage is to continue with some further comparison. That way, the DM's responses can be checked for consistency and greater accuracy can evolve. Even though the Singapore study demonstrated that complete comparisons were more accurate than $n-1$ or sufficient comparisons, there is no proof that the final priorities were the most accurate of all the possible priorities. An interesting question that still requires testing is whether some level between sufficient and complete is more accurate.

Determining the optimal number of comparisons is a multi-criteria problem. There are multiple objectives (accuracy, ease of use, time, etc) and how each DM trades off these criteria will vary. Ultimately, it is the DM who chooses the priorities that best represent his or her preference. If the DM stops before complete comparisons, then those must be the priorities. Nobody else can determine the DM's preferences.

8. Conclusion

The major advantage of the 4 step process is to achieve efficiency without unnecessarily compromising accuracy. The DM has the option to complete all comparisons or to stop after having performed as few as $n-1$ comparisons. The recommended process is to undertake a sufficient number of additional comparisons beyond $n-1$. Even though those additional comparisons are redundant to calculating priorities, they provide evidence that the DM is consistent and logical in his/her evaluations. The DM can accept results from incomplete comparisons with assurance that they are close to what would be achieved with complete comparisons.

The advantages of the 4 step process go beyond fewer comparisons and greater efficiency. The initial ranking structures the problem in a manner that keeps all items before the evaluator. If ordinal errors occur, that can be easily visualized. If ranks are incorrect, that too can be easily visualized. Since all intensities are ranked and seen at one time, it is quite possible that accuracy will be enhanced. The 4 stage process warrants research into its efficacy. As well, it could be implemented in AHP/ANP computer programs.

REFERENCES

- Carmone, F.J., Kara, A. & Zanakis, S H. (1997). A Monte Carlo investigation of incomplete pairwise comparisons matrices in AHP, *European Journal of Operational Research*, 102, 538-553.
- Fedrizzi M. & Giove. S. (2007). Incomplete pairwise comparison and consistency optimization, *European Journal of Operational Research*, 183, 303–313
- Harker, P. T. (1987a). Shortening the comparison process in the AHP. *Mathematical Modelling*, 8, 139-141
- Harker, P. T. (1987b). Alternative modes of questioning in the analytic hierarchy process *Mathematical Modelling*, 9(3-5), 353-360.
- Harker, P. T. (1987c). Incomplete pairwise comparison in the analytic hierarchy process, *Mathematical Modelling* 9(11), 837-848.

- Huizingh E. & Vrolijk H. (1997). A comparison of verbal and numerical judgments in the Analytic Hierarchy Process, *Organizational Behaviour and Human Decision Processes*, 70(3), 237-247
- Ishizaka A. & Lusti, M. (2004) An expert module to improve the consistency of AHP matrices, *International Transactions in Operational Research*, 11, 97-105.
- Michell, J. (1997). Quantitative science and the definition of measurement in psychology, *British Journal of Psychology*, 88, 355-383.
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information, *Psychological Review*, 63(2), 81-97.
- Millet I. & Harker, P. T. (1990). Globally effective questioning in the Analytic Hierarchy Process, *European Journal of Operational Research*, 5, 88-97.
- Narens, L. (1996) A theory of ratio magnitude estimation, *Journal of Mathematical Psychology*, 40, 109-129.
- Setiawan, S. (1997). The effects of initial selections in estimating the missing comparisons in an incomplete matrix, MSc Thesis, Department of Industrial and Manufacturing Systems Engineering, Louisiana State University.
- Shen, Y., Hoerl, A.E. & McConnell, W. (1992) An incomplete design in the analytic hierarchy process, *Mathematical and Computer Modelling*, 16(5), 121-129.
- Saaty, T.L. (1977). A scaling method for priorities in hierarchical structures, *Journal of Mathematical Psychology*, 15, 234-281.
- Saaty T. L. & Ozdemir, M. S. (2003) .Why the magic number seven plus or minus two, *Mathematical and Computer Modelling*, 38(3-4), 233-244.
- Stevens, S. S. (1946), "On the theory of scales of measurement", *Science*, 103, 677-680
- Stevens, S.S. (1971). Issues in psychophysical measurement, *Psychological Review*, 78(5), 426-450.
- Takeda, E. & Yu, P. L. (1988). Eliciting the relative weights from incomplete reciprocal matrices. *Preprints of the International Symposium on the Analytic Hierarchy Process*, Tianjin, China, September 6-9, 192-200.
- Webber, S. A., Apostolou, B. & Hassell, J. M (1996). The sensitivity of the analytic hierarchy process to alternative scale and cue presentations. *European Journal of Operational Research*, 96, 351-362.
- Wedley, W. C. (1993). Consistency prediction for incomplete AHP matrices, *Mathematical Modelling*, 17(4/5), 151-161.
- Wedley, W. C., Schoner, B. & Tang, T. S. (1993). Starting rules for incomplete comparisons in the Analytic Hierarchy Process, *Mathematical Modelling*, 17(4/5), 151-161.
- Whitaker, R. (2007) .Validation examples of the Analytic Hierarchy Process and Analytic Network Process, *Mathematical and Computer Modelling*, 46, 840-859.