

Using the Analytical Hierarchy Process to Evaluate Target Signatures

J Baumbach

CSIR, Defence Peace Safety Security (DPSS), South Africa

ABSTRACT

The Analytical Hierarchy Process (AHP) and the Law of Comparative Judgement (LCJ) are pairwise comparison methods. A large number of observers need to perform an LCJ evaluation in order to get accurate results. LCJ also does not provide an absolute scale of performance, nor does it provide a metric to evaluate the accuracy of the evaluation. Abovementioned shortcomings were addressed using AHP.

The AHP was used to evaluate the effectiveness of camouflage patterns printed on fabric. The camouflage uniforms were presented to a panel of observers two at a time. The observers scored the effectiveness according to a scale provided. This data was then used to calculate the relative effectiveness values in order to rank the patterns from the best to the worst performer. AHP also allows for a metric to indicate how consistent an observer assigned scores to the different uniforms.

1. Introduction

One of the aspects of camouflaging is the effective usage of colours within patterns, which allows an object to blend with the environment. Refinement of colours and patterns involves, amongst others, live field trials. Various methods are currently used to qualify and quantify comparisons between different systems, of which the most popular are: the Law of Comparative Judgment (LCJ) [1, 2], a sliding scale method as used by Dugas [3] and the calculation of the cumulative probability of detection [4].

The Analytical Hierarchy Process (AHP), a pairwise comparison method developed by Saaty [5], is widely used in the commercial environment as a decision support tool for selecting between a number of alternative

options. As far as we know AHP has never been used as an evaluation method for camouflage systems. We have adopted AHP to be used for this purpose, specifically where multiple uniforms need to be compared for their suitability (in terms of pattern and colour) in a variety of environments. We will compare the results from AHP with those of LCJ, which is also a pairwise comparison method.

2. Analytical Hierarchy Process

AHP give priority weights to stimuli [5, 6], something the other methods mentioned does not do. This ensure much more accurate ranking of a number of stimuli.

Observers have to state how much better/heavier/louder stimulus A is,

compared to stimulus B. The “how much” is then rated according to a scale, as shown in .

Table 1. Comparison Scale

Scale	Definition
1	As good as
3	Marginally better
5	Much better
7	A lot better
9	Extremely better
2,4,6,8	Intermediate values

A number of observers (n) then evaluate the stimuli, which is presented two at a time to them. For each observer the matrix “A” is completed, using the scale provided.

For every value in the matrix given to any of the comparisons (e.g. stimuli A vs. B), the B-A comparison gets the reciprocal value. The diagonal of the matrix is unity. Therefore, the A-matrix looks as follows:

$$A_{AHP} = \begin{bmatrix} 1 & a_{12} & \cdots & a_{1n} \\ 1/a_{12} & 1 & \cdots & a_{2n} \\ \cdots & \cdots & 1 & \cdots \\ 1/a_{1n} & 1/a_{2n} & \cdots & 1 \end{bmatrix}$$

The A-matrix can be completed for every observer, or the mean for all the observer’s scores can be calculated to be combined in a single matrix. When the results for a number of observers are combined, the geometric mean for each element over all observers is calculated before populating the A-matrix [5, 9]. In this case, each element is calculated as follows:

$$a_{ij} = \left(\prod_{i=1}^n a_i \right)^{1/n} = \sqrt[n]{a_1 \cdot a_2 \cdots a_n}$$

Next, the largest eigenvalue (λ_{\max}) is calculated, together with its associated eigenvector (w):

$$Aw = \lambda_{\max} w$$

The weights are calculated by normalising the eigenvector:

$$Weights = \frac{w}{\sum_{i=1}^n w_i} * 100$$

A consistency index is defined, which indicate how consistent the comparisons were made. The consistency index (CI) is defined as:

$$CI = \frac{(\lambda_{\max} - n)}{(n-1)}$$

After this a consistency ratio (CR) is calculated, indicating how consistent the evaluation is performed, relative to the average of a large number of matrixes populated with random generated numbers (called the Random Consistency Index (RCI), as published by Saaty [5]). The CR is calculated as follows:

$$CR = \frac{CI}{RCI}$$

If values of less than 0.1 (10%) for the CR are obtained, it is considered as a very consistent evaluation. Values between 10% and 20% imply acceptable consistency. Depending on requirements it is advisable to repeat the evaluation for values more than 10%. There is also a relationship between λ_{\max} and the number of stimuli: the closer the largest eigenvalue is to the number of stimuli, the more accurate the evaluation.

3. Experimental Design

This method was specifically used for camouflage pattern evaluations. Since it is a pairwise comparison method, the uniforms were presented to the observers two at a time (Figure 1).



Figure 1 Camouflage pattern evaluation

The observers were forced to make a choice between one of the two uniforms, by asking the question: “which one of the two uniforms presented to you would you prefer to wear if you need to hide from enemy observation in this environment?”. The observers then need to assign a score to their choice, through the question: “how much better is your chosen uniform, if compared to the other one presented?”.

All of the observers’ scores are captured on a recording sheet.

It is important to note that the more observers performing the evaluation, the better the probability of having a larger number of consistent evaluations. Also, the number of observations increases with the square of the number of uniforms; therefore through our experience we advise to have not more than five uniforms during a single evaluation.

4. Data Analyses: AHP

The scores for all the observers were captured in a spreadsheet. Custom Matlab code was used to read the data, and calculate the weights and the CR. The weights, the CR and the eigenvalue for each uniform, as evaluated by the observers, are shown in the left-hand side of Table 2. This is shown under the heading “AHP (weights for each observer)”.

Thereafter, three different options for combining the scores of all the observers were investigated. These three options are shown in Table 2, under the heading “AHP (averaged weights)”.

The first option was to calculate the geometric mean of all the observer’s data (that is all data with $CR > 0$). The resulting A-matrix was then used to calculate the weights and the CR for the combined data. Also, the standard deviation of the calculated weights for all observers, but for each uniform, was calculated. These values are shown in Table 2.

The second option was to calculate the mean using only the data for observers which scored the uniforms consistently (that is observers with a $CR < 20$). The weights and CR was then calculated. As above, the standard deviation in the weights for each uniform was calculated, but only the data for observers with $CR < 20$ were used. These values are shown in the appropriate column of Table 2.

The last option, which was only done for illustrative purposes, was to calculate the geometric mean of the data for observers which scored inconsistently (observers with $CR > 20$). The resulting standard deviation, together with the weights and CR are also shown in Table 2.

5. Data Analyses: LCJ

The Law of Comparative Judgement (LCJ) was developed by Thurstone in 1927 [7, 8]. This is also a pairwise comparison method. In contrast to AHP, a large number of observers (n) are needed to get accurate results, because LCJ assumes a normal distribution for all evaluations. Only the number of times a specific uniform is selected is noted, e.g. if all six observers prefer Pattern1 over Pattern2, the numbers “6” and “0” is used in the relevant calculations.

The data of the evaluation as discussed earlier was also analysed using the LCJ method. The scores for each of the patterns are shown in the last column of Table 2.

The LCJ weights are different to the AHP weights in the sense that the lower the score value, the better the performance.

Table 2.AHP and LCJ Results

		AHP (weights for each observer)						AHP (averaged weights)						LCJ
								CR > 0		CR < 20		CR > 20		
		Observer1	Observer2	Observer3	Observer4	Observer5	Observer6	Rank	Std Dev	Rank	Std Dev	Rank	Std Dev	6 Observers
Weights	Pattern1	63	50	51	62	56	49	58	6.2	53	3.2	62	7.8	-2.27
	Pattern2	4	12	7	24	8	14	10	7.1	9	2.6	12	10.0	0.60
	Pattern3	13	9	16	8	9	8	11	3.3	11	4.0	10	2.9	0.92
	Pattern4	20	29	26	5	26	29	21	9.2	27	1.7	16	12.1	0.64
CR		38	19	8	31	4	29	4		5		3		
λ_{max}		5.00	4.50	4.13	4.83	4.11	4.77	4.10		4.13		4.08		

6. Discussion

With visual inspection of the AHP weights it is obvious that all observers place Pattern1 in the first position, followed by Pattern4. There is no obvious rank preference for Pattern2 and Pattern3.

The CR for three of the observers was less than 20, meaning they scored the uniforms' performance consistently. Also note that the smaller the CR, the closer the eigenvalue (λ_{max}) is to number of alternatives (four in this case). When these observers's data was combined using the geometric mean, the CR decreased to 5%, which is regarded as very consistent.

The standard deviation also increased dramatically for the case where the data for the observers with inconsistent scoring were combined (observers with CR>20). However, the CR is 3, when the weights for this dataset are calculated. This is due to using the geometric mean to calculate the average. When making decisions regarding the consistency, it is important not to look

at the value in isolation, but in relation to the standard deviation.

When looking at the standard deviation for both the cases where CR>0 and CR>20, none of them have a conclusive rank for the last three uniforms. In the case where CR<20, the rank for only the last two uniforms was inconclusive.

The LCJ score reveals that Pattern1 is evaluated to be the best. The scores for the other three patterns are very close, but this method gives no indication of how significant these differences are.

For comparison, the AHP weights (CR<20) and the LCJ scores are plotted on the same graph (Figure 2). To be consistent in the graphical presentation, the LCJ score was multiplied with -1, so that the best score is represented by the largest number.

From the graph it is evident that Pattern1 is by far the most effective in this environment, followed by Pattern4, as indicated by AHP. AHP indicates the effectiveness of Pattern2 and Pattern3 are almost the same. On the contrary, LCJ

indicate the effectiveness of Pattern2, -3 and -4 to be almost the same.

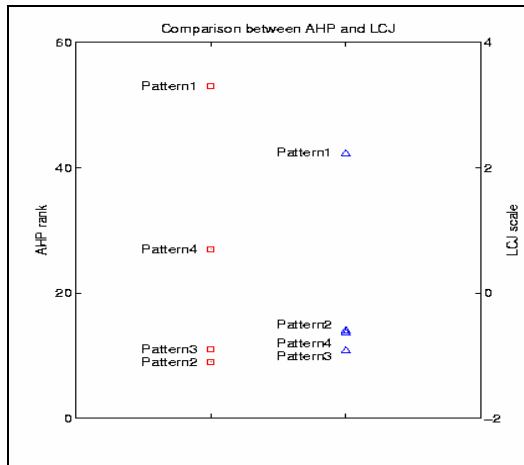


Figure 2 AHP and LCJ Comparison

When using AHP, two very common types of evaluation errors are mentioned in the literature [10]. These are:

- Type 1: Consistent judgment, inaccurate priorities. This happens when the CR is low, but the priorities are not accurate.
- Type 2: Inconsistent judgment, accurate priorities. This is when the CR is unacceptably high, but the priorities are accurate.

We have found that Type 1 errors could either be attributed to an observer's bias towards a certain pattern, or to a possible vision problem (e.g. the observer is not wearing his prescription glasses).

Type 2 errors are the most common (that is when $CR > 20$). This occurs when the observers do not assign realistic scores during the evaluation.

When the scores of the observers are captured, and a certain comparison cannot be made, AHP allows zero's in the appropriate positions of the A-matrix.

However, when calculating the geometric mean of several observers' scores, a zero at any position immediately nullify that specific data point for all the observers. A way around this problem would be to omit the zero entry, and instead of taking the n-th root, using the (n-1)th root. We have only worked with complete matrices, and have not explored the effect of such a calculation on the weights.

7. Conclusions

AHP and LCJ are pairwise comparison methods used to evaluate camouflage effectiveness, but we regard AHP to give more meaningful results than LCJ.

AHP give a relative performance metric between multiple samples, i.e. the question of "how much better" can be answered. It uses an absolute, linear scale (0 to 100). This means that if the weight of X is 10 and Y is 20, it means Y is two times more effective than X. Another advantage is that it does not require as many observers as with LCJ in order to give accurate weights for the different patterns.

However, it becomes very time consuming when a large number of alternatives are evaluated. This is the reason why a maximum of five alternatives during a single evaluation is mentioned earlier.

We have not tested any correlation between AHP and the probability of detection, as used by NATO [4]. This is mainly because we do not have the necessary facilities to execute such a test.

Our data collection and analyses with AHP was only performed with live trials in the field. However, it should be possible to use it with photo simulation trials as well. This will include projection of the imagery on a screen (computer screen or data projector) as well as printed photographs.

8. References

- [1] J.R. McManamey, *Applying the Law of Comparative Judgement to Target Signature Evaluation*, RTO CSI Workshop on "Search and Target Acquisition", Utrecht, The Netherlands, 21-23 June 1999
- [2] J.J. Crile, J.R. McManamey, *Validation of a New Analyses Program for the Law of Comparative Judgement Data*, 13th Ground Target Modelling and Validation Conference, Houghton, MI, 5-8 August 2002
- [3] A. Dugas, F.M. Kramer, *Universal Camouflage for the Future Warrior*, International Soldier Systems Conference and Exhibition, 15 December 2004
- [4] J.E. Peak, e.a., *Guidelines for Camouflage Assessment Using Observers*, Research and Technology Organisation (RTO), North Atlantic Treaty Organisation (NATO), RTO-AG-SCI-095, AC/323(SCI-095)TP/96, October 2006
- [5] T.L. Saaty, *The Analytic Hierarchy Process*, McGraw-Hill, New York, 1980
- [6] T.L. Saaty, *How to make a Decision: The Analytic Hierarchy Process*, Interfaces, Volume 24(6), 1994, pp 19-43
- [7] L.L. Thurstone, *A Law of Comparative Judgment*, Psychological Review, Vol 34, 1927, pp 273-286
- [8] W.S. Torgerson, *Theory and Methods of Scaling*, John Wiley & Sons, New York, 1958
- [9] J. Aczel, T.L. Saaty, *Procedures for Synthesizing Ratio Judgments*, Journal of Mathematical Psychology, Volume 27, 1983, pp 93-102
- [10] W.C. Wedley, *Consistency Prediction for Incomplete AHP Matrices*, Mathematical and Computing Modeling, 17/4-5, 1993, pp 151-162