RELATIVE MEASUREMENT OF THE CONSISTENCY RATIO

William C. Wedley
Faculty of Business Administration
Simon Fraser University
Burnaby, B. C.
CANADA, V5A 1S6

ABSTRACT

Since it provides varying degrees of discrimination across matrix sizes, the 10 percent cut-off requirement for the Consistency Ratio (CR) is, in itself, an inconsistent rule. By looking at the CR as a relative departure from pure consistency (rather than pure randomness), this study proposes a better measure. It simulates varying degrees of departure from pure consistency, and it analyzes two different consistency measures. It concludes by recommending advice for AHP users depending upon their relative departure from pure consistency.

INTRODUCTION

In the Analytic Hierarchy Process (AHP), the Consistency Ratio has proved to be very helpful to decision makers in deciding whether they have been logical in entering their paired comparisons. Thomas L. Saaty (1977, 1980), the originator of AHP, defined this measure as being the ratio of a consistency index of a matrix to the mean consistency index from a large sample of randomly generated matrices.

$$\text{Consistency Ratio (CR)} = \text{CI} / \text{Mean Random CI} \tag{1}$$

Central to the measurement of CR is the consistency index.

$$\text{Consistency Index (CI)} = \frac{\lambda_{max} - 1}{n - 1} \tag{2}$$

In this formula, λ_{max} is the largest principal eigenvalue of the positive reciprocal pairwise comparison matrix ($a_{ij}=1/a_{ji}$) of size n . If the paired comparisons are perfectly consistent ($a_{ij} * a_{jk} = a_{ik}$), then λ_{max} is equal to the size of the matrix and the consistency index is zero. The larger the inconsistency between comparisons, the larger λ_{max} and the larger the consistency index (which really measures inconsistency). In matrices with random entries, CI is likely to be very large.

By placing the Mean Random CI in the denominator of (1), Saaty defines the CR as the degree of departure from pure inconsistency. If a person has been relatively consistent in making their paired comparisons, then they should have a much lower consistency index than what would be produced by random entries. Saaty claims that an acceptable consistency ratio should be less than .10, although a ratio of less than .2 is considered tolerable.

Although this rule has provided good guidance, it, in itself, has been found to be inconsistent across matrix sizes (Lane and Verdini, 1989). Table 1 illustrates this fact. It shows the mean random consistency index for different matrix sizes plus the consistency index at various percentiles on

Table 1. Random and Cutoff Consistency Indexes

Mean Size of Matrix	Consistency Index from Randomly Generated Matrices n = 500	Consistency Index at Various Percentiles				Saaty's Cut-off Consisteny Indexes	
		20%	10%	5%	2%	Tolerable Level 20%	Recommended Level 10%
3	.58	.027	.005	.001	.000	.116	.058
4	.90	.237	.130	.075	.035	.180	.090
5	1.12	.564	.392	.302	.217	.224	.112
6	1.24	.820	.619	.495	.342	.248	.124
7	1.32	.951	.790	.648	.466	.264	.132
8	1.41	1.073	.924	.839	.698	.282	.141
9	1.45	1.155	1.022	.908	.837	.290	.145
10	1.49	1.191	1.096	1.023	.892	.298	.149

the consistent side of the distribution. Also included are Saaty's recommended cut-off points for acceptable consistency. Notice that for matrix sizes 5 or larger, the cut-off consistency index is well outside the randomly generated distribution. For matrix sizes 3 and 4 there is a greater than 5 percent chance that a consistency index at the cut-off point could have come from the distribution of random entries.

The implication is that perhaps the 10 percent cut-off is too lax for the smaller-sized matrices but too onerous for larger matrices. This is logical, because larger-sized matrices have more redundant comparisons and therefore greater opportunities for inconsistencies to creep into the judgments. Perhaps a more sensible policy would be a variable cut-off (not 10 percent) which increases with the size of matrix.

Lane and Verdini (1989) made a partial step in this direction when they suggested stricter cut-off requirements for 3 and 4 attribute matrices (.0035 and .048 respectively). For larger-sized matrices, they would maintain Saaty's 10 percent rule. One wonders, however, whether the .0035 level is achievable for matrix size 3 and whether the 10 percent rule is appropriate for larger-sized matrices.

An alternate approach adopted by Golden and Wang (1989) was to develop an entirely new measure of consistency. Their measure is

$$G = 1/n \sum \sum \left| C^*_{ij} - g^*_i \right| \qquad (3)$$

where C^*_{ij} is the column normalized paired comparison matrix
g^*_i is the normalized vector of row geometric mean priorities

Their method also works with the traditional eigenvector priorities (e^*_i) in place of g^*_i. With perfect consistency, all column normalized vectors of C^*_{ij} are equal to g^*_i or e^*_i and the consistency measure would be equal to zero.

To specify cutoffs for their new measure, Golden and Wang produced samples of simulated matrices with varying degrees of inconsistency. They showed that the 33rd percentile of the sample with perturbations of ±3 intervals on the AHP rating scale produced results similar to Saaty's 10 percent rule, yet with more uniform discrimination across different matrix sizes

While it appears that Golden and Wang's cut-off values are an improvement over the 10 percent rule, it is not as clear that a new measure is necessary. As alluded above, a variable cut-off for the CI or CR (not a set 10 percent) could also provide more uniform discrimination.

This study investigates the use of a different CR and compares it to Golden and Wang's G value. Unlike the traditional CR which measures departure from pure randomness, this study looks upon the inconsistency measure as a departure from pure consistency. In other words, the CR proposed herein utilizes a maximum acceptable CI in the denominator of equation (1). A value of greater than one would then suggest a level of inconsistency beyond the acceptable level.

In part, this study is similar to Golden and Wang's approach. They too, looked upon inconsistency as a departure from the purely consistent case. It differs, however, in the manner in which the simulations and validations are conducted. In their study, n-1 comparisons from the first row of the paired comparison matrix were randomly selected from the AHP rating scale (1/9, 1/8, ... 1/3, 1/2, 1, 2, 3, ... 8, 9). These values were then used to generate temporarily consistent comparisons for the remaining $n(n-1)/2-(n-1)$ items. These temporary items were then perturbated ±k intervals on the AHP rating scale , subject to the constraint that $1/9 \le C_{ij} \le 9$. Since the first n-1 comparison are not perturbated and since each matrix size has a different number of perturbated comparisons, this procedure produces a different average perturbation across matrix sizes for any selected k. Secondly, the calculation of the consistent (before perturbation) comparisons ($C_{ij} = C_{1j}/C_{1i}$) can result in values far in excess of what is allowed in AHP theory. This necessitates excessive use of the $1/9 \le C_{ij} \le 9$ constraint to bring the values back within range. The procedure proposed herein helps overcome these deficiencies.

METHOD

The premise used in this study is that measurement of the degree of inconsistency should be considered as a departure from pure consistency. Accordingly, perfectly consistent matrices (n=500) were simulated and these were subjected to 10 different levels of perturbation. The perfectly consistent matrices and their perturbations were conducted in the following manner.

Generating the Perfectly Consistent Matrix

1. Generate a Random Number N(1) between 10 and 100 and initially designate this number as both the Minimum (N_{min}) and Maximum (N_{max}) of n numbers to be generated.

2. For i= 2 to n,
   a. determine N_{min} and N_{max}.
   b. randomly select N(i) from the range 1/9•(N_{max}) and 9•(N_{min}).

3. The perfectly consistent paired comparisons are A(i,j) = N(i)/N(j)

   where A(i,j) is the perfectly consistent paired comparison matrix.

Perturbing the Perfectly Consistent Matrix

For all A(i,j) ≥ 1.0

1. First Perturbation

   Randomly change each paired comparison A(i,j) to the next higher or lower integer P(i,j) (not the same as rounding). P(j,i) = 1 / P(i,j)

2 Perturbations 2 to 10

   For d = 1 to 9

   $$
   \begin{array}{lll}
   P(i,j) = A(i,j) + k & \text{if } 1.0 \le A(i,j) + k \le 9.9 & \\
   P(i,j) = 9.9 & \text{if } A(i,j) + k > 9.9 & (k \text{ is } +) \\
   P(i,j) = 1/(|\,k\,| - A(i,j) + 2) & \text{if } A(i,j) + k < 1.0 & (k \text{ is } -)
   \end{array}
   $$

   where P(i,j) is the perturbated paired comparison
   k is a random ± integer from 1 to d.

   P(j,i) = 1 / P(i,j)

In total, there are ten different degrees of perturbation. In the first perturbation above, the raising or lowering of each comparison (A(i,j) ≥ 1.0) to the higher or lower integer results in a theoretical average perturbation value of .5 interval step on the AHP rating scale. By design, the other d perturbation levels will have theoretical average perturbation values of

d =	1	2	3	4	5	6	7	8	9
average perturbation	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0

For example, with d = 3, the potential values for k are -3, -2, -1, 1, 2, and 3. The absolute average of these is 2.0. In practice, however, the actual average will generally be lower than these values, because the truncation of extreme P(i,j) to 9.9 causes the effect of some perturbations to be dampened. In passing, we should note that the last 9 perturbations wit d= 1 to 9 allow the P(i,j) to be in non-integer form, just like the original A(i,j).

For each perturbated matrix, we can calculate the average perturbation value and the resulting CI and G. We can use these data to determine appropriate cut-off points and we can also use them with a separate sample to see how each method discriminates between consistent and inconsistent cases.

SIMULATION RESULTS

The mean CI and G values for various matrix sizes and levels of perturbation are presented in Tables 2 and 3 respectively. Notice first that the average perturbation value of the various matrix sizes departs from the theoretical average as the degree of perturbation increases. This is because larger-sized perturbations are truncated to 9.9 if they go above this value.

Table 2. Statistics for CI at Different Perturbation Levels

Matrix Size and Type of Measure	Range (±) of Non-Zero Integer Perturbations (k) and the Range of Resulting Average Perturbation Intervals									
	N. I .48-.49	± 1 1.00	± 2 1.49-1.50	± 3 1.98-2.00	± 4 2.45-2.48	± 5 2.93-2.95	± 6 3.37-3.44	± 7 3.85-3.87	± 8 4.20-4.32	± 9 4.64-4.76
3 Mean CI	.023***	.034***	.069***	.117***	.182***	.222***	.298***	.355***	.455***	.457***
Std. Deviation	.037	.034	.089	.149	.241	.296	.409	.515	.624	.660
4 Mean CI	.040***	.058***	.121***	.205***	.298***	.418***	.530***	.665***	.810***	.902***
Std. Deviation	.040	.034	.087	.162	.228	.325	.405	.489	.578	.662
5 Mean CI	.050***	.075**	.164	.265***	.403**	.527*	.681**	.809	1.002*	1.126*
Std. Deviation	.035	.032	.079	.137	.210	.273	.361	.408	.498	.555
6 Mean CI	.059***	.091	.195	.323	.476	.642	.800	.974	1.158	1.309
Std. Deviation	.034	.028	.070	.120	.180	.235	.308	.341	.403	.455
7 Mean CI	.064**	.097	.217	.354	.527	.683	.871	1.069	1.245	1.408
Std. Deviation	.029	.026	.061	.108	.158	.198	.239	.282	.343	.376
8 Mean CI	.071*	.106	.232	.381	.551	.728	.929	1.119	1.285	1.498
Std. Deviation	.028	.023	.057	.096	.133	.178	.213	.252	.283	.306
9 Mean CI	.079	.111	.245	.399	.592	.784	.964	1.170	1.371	1.573
Std. Deviation	.027	.022	.052	.086	.117	.144	.181	.204	.243	.274
10 Mean CI	.081	.117	.256	.422	.600	.810	1.005	1.208	1.412	1.601
Std. Deviation	.024	.020	.047	.073	.108	.137	.170	.186	.208	.239

Note N. I. = Next Integer CI = Consistency Index
Kolmogorov-Smirnov test for normality *** = Z > 1.94, p<.001
** = Z > 1.62, p<.01
* = Z > 1.36, p<.05

Table 3 Statistics for G at Different Perturbation Levels

Matrix Size and Type of Measure	Range (±) of Non-Zero Integer Perturbations (k) and the Range of Resulting Average Perturbation Intervals									
	N. I .48-.49	± 1 1.00	± 2 1.49-1.50	± 3 1.98-2.00	± 4 2.45-2.48	± 5 2.93-2.95	± 6 3.37-3.44	± 7 3.85-3.87	± 8 4.20-4.32	± 9 4.64-4.76
3 Mean G	.101***	.142	.187**	.241***	.287***	.308***	.337***	.361***	.399***	.388***
Std. Deviation	.090	.082	.127	.160	.195	.212	.246	.255	.284	.283
4 Mean G	.161**	.207***	.283*	.353**	.397	.461*	.495***	.540**	.578***	.592**
Std. Deviation	.086	.072	.111	.144	.160	.180	.196	.211	.222	.217
5 Mean G	.192	.247**	.346	.416	.491	.541	.589*	.620*	.660*	.682**
Std. Deviation	.077	.065	.094	.112	.128	.144	.154	.153	.162	.166
6 Mean G	.216	.289	.392	.479	.551	.609*	.653	.692**	.729**	.749*
Std. Deviation	.067	.060	.082	.096	.110	.111	.121	.123	.124	.123
7 Mean G	.232	.306	.424	.511	.592*	.638	.690*	.736	.765	.787*
Std. Deviation	.057	.053	.071	.085	.091	.091	.097	.094	.095	.097
8 Mean G	.248	.325	.448	.537	.614	.668*	.719*	.759*	.786*	.813*
Std. Deviation	.054	.049	.063	.072	.076	.082	.078	.079	.081	.078
9 Mean G	.263	.339	.465	.561	.638	.697	.740	.783	.809	.839*
Std. Deviation	.050	.046	.056	.064	.067	.064	.065	.064	.065	.064
10 Mean G	.270	.355	.481	.579	.652*	.718	.758	.797	.828	.852
Std. Deviation	.046	.043	.054	.056	.060	.058	.170	.055	.051	.053

Note N. I. = Next Integer G = Golden and Wang's G value
Kolmogorov-Smirnov test for normality *** = $Z > 1.94$, $p<.001$
** = $Z > 1.62$, $p<.01$
* = $Z > 1.36$, $p<.05$

As expected, both CI and G increase in a fairly regular fashion as the degree of perturbation increases. As one can see by a comparison with Table 1, when $k = \pm 9$, the simulated CI is very close to the results from the random entry of comparisons. Also by comparison to Table 1, we can see that Saaty's 10 percent cut-off falls between $k = \pm 1$ and $k = \pm 2$ and an average perturbation value of between 1.0 and 1.5.

The 33rd percentiles which Golden and Wang use as cut-offs are .103, .196, .259, .299, .323, .344, .361, .371 for matrices 3 to 10 respectively. On Table 3, these fall between the average perturbation intervals of .48 and 1.00 for matrix sizes 3 and 4 and between average perturbation levels of 1.0 and 1.49 for matrix sizes 5 to 10.

Except for the random changes to the next higher or lower integer, the distributions are approximately normal for $n \geq 6$ for CI and $n \geq 5$ for G. Although space does not allow presentation of the actual Kolmogorov-Smirnov Z-values, the detailed data illustrate that G has a more uniform distribution, particularly with $n<7$. At higher levels of perturbation and with $n \geq 7$, CI comes closer to approximating the normal distribution.

A NEW CONSISTENCY RATIO

In order to determine new cut-offs which measure relative departure from consistency, it is necessary to pick an upper inconsistency level which is regarded to be unacceptable. In this study, we have chosen the distribution with an average perturbation value of 1.5 (column with $k=\pm 2$ in Table 2). These mean CI values are greater than Saaty's 10 percent cut-off rule but lower than 20 percent. They are also well the comparable 33 percent cut-off for G.

With this as a new upper level, two new types of cut-offs can be developed for CI and G. A new consistency ratio can be defined as

$$\text{Consistency Ratio (CRm)} = \text{CI} \,/\, \text{Mean CI}_{(k=\pm 2)} \tag{4}$$

With this measure, an unacceptable CRm will have a value greater than unity. Moreover, values less than unity can signify differing degrees of consistency and we can use give more specific advice rather than a dichotomous acceptable/unacceptable statement. For example, the N. I. perturbation column of Table 2 has a CRm between .29 and .33, while the $k=\pm 1$ column has CRm between .45 and .49. Thus appropriate advice might be

CRm < .31	Excellent Consistency
.31 ≤ CRm < .47	Good Consistency
.47 ≤ CRm < 1.0	Acceptable Consistency but Caution Needed
CRm ≥ 1.0	Unacceptable Inconsistency

From Table 3, the same type of ratio and advice can be developed for G except that the cut-off ratios would be .55, .73 and 1.

A second type of cut-off is the break point which results in an equal percentage of observations in each overlapping distribution. For both CI and G, these values for distributions N. I to +3 are presented in Table 4. Notice that there is considerable overlap between the various distributions. This is caused by the small gradations of perturbation between distributions. Although the overlap is severe for both G and CI, there is higher overlap for the G measure.

These break points can be used as statistical tests for deciding which distribution a generated CI or G comes from. As was done for the means of the distributions, they can also be used to establish a consistency ratio which allows variable advice.

$$\text{Consistency Ratio (CRbp)} = \text{CI} \,/\, \text{Break Point CI}_{(k=\pm 2,\, vs.\, k=\pm 3)} \tag{5}$$

From Table 4, the variable advice cut-off ratios would be .31, .52, and 1. for CI, and .58, .77, and 1. if G is used in the above equation rather than CI.

Table 4 CI and G values for Break Points Between Distributions

Matrix Size	Break Points and Percentage Overlaps in Distributions NI - ±3											
	BPe1 NI vs. ±1		BPe2 +1 vs. ±2		BPe3 ±2 vs. ±3		BPg1 NI vs. ±1		BPg2 ±1 vs. ±2		BPg3 ±2 vs. ±3	
	CI	%	CI	%	CI	%	G	%	G	%	G	%
3	.017	(36)	.030	(32)	.052	(42)	.111	(37)	.152	(43)	.189	(43)
4	.039	(36)	.068	(32)	.120	(38)	.177	(37)	.225	(32)	.294	(40)
5	.057	(34)	.100	(24)	.187	(35)	.215	(38)	.284	(29)	.375	(38)
6	.074	(28)	.120	(14)	.235	(26)	.253	(29)	.332	(25)	.434	(32)
7	.080	(26)	.133	(9)	.263	(23)	.270	(27)	.354	(19)	.463	(31)
8	.089	(24)	.143	(6)	.288	(18)	.287	(24)	.374	(16)	.490	(25)
9	.097	(25)	.155	(3)	.302	(14)	.302	(24)	.395	(11)	.514	(23)
10	.100	(21)	.157	(3)	.316	(8)	.311	(19)	.410	(10)	.528	(18)

VALIDATION

The samples for comparing and validating the two consistency measures were simulated from the perturbation levels represented by the first 4 data columns of Tables 2 and 3 (i. e. from N. I. to k=±3). In other words, only perturbations of ±3 causing average perturbation values of ≤2.00 were considered within the realm of reasonable levels which people would generate in practice. For each matrix size, 250 observations were taken from each of the 4 perturbation levels (n=1000).

This simulation procedure is actually an amalgamation of 4 overlapping distributions where the average paired comparison perturbation is 1.24 (i. e. the average of the perturbation levels listed in the first 4 columns of Table 2 or 3). While this procedure maintains comparable perturbation levels across matrix size, it also causes the validation samples to be dominated by low CI or G and to be skewed to the right. Accordingly, all validation distributions of CI and G are non-normal (minimum Kolmogorov-Smirnov Z=1.73, p<.005 for G; Z=4.42, p<.000 for CI).

Table 5 summarizes the acceptance rates for various types of cut-offs. Presented is Saaty's 10% rule for the CR, Lane and Verdini's modified 10% rule, Golden and Wang's 33 percentile for G, and the variable cut-off ratios for CI and G proposed in this study. Also included are the Golden and Wang results from their validation with k=3.

The first point to note is that the Saaty's 10 percent rule is very stable in its discriminatory performance for n>5. For smaller-sized matrices, where skewness is more extreme, it is too lenient in concluding that sufficient consistency is present. Lane and Verdini's modification to to use a more restrictive acceptance criteria for n=3 and n=4 overcompensates, because few matrices are accepted at those levels and the standard deviation of the mean acceptance rate increases from 71 to 111. Golden and Wang's 33 percentile, with a standard deviation of 16, displays consistent discrimination across matrix sizes. By rejecting more matrices, it is also a more restrictive rule.

Table 5 -- Acceptance Numbers for Various Cutoff Criteria (n=1000)

Measure of Effectiveness	Matrix Size and Average Perturbation Size								Std Dev
	3 ±1.24	4 ±1.25	5 ±1.25	6 ±1.24	7 ±1.24	8 ±1.25	9 ±1.25	10 ±1.24	
CR < 10%	693	601	551	493	490	489	487	486	71
Modified CR < 10%	173	382	551	493	490	489	487	486	111
G < 33rd percentile	373	400	403	419	416	426	417	415	16
Ratios from Means									
CRme < .31	430	299	231	173	177	152	135	130	96
CRmg < .55	371	250	192	134	147	135	115	111	83
CRme < .47	**558**	**427**	**385**	**344**	**374**	**390**	**351**	**383**	**64**
CRmg < .73	**489**	**428**	**383**	**366**	**375**	**387**	**364**	**367**	**40**
CRme < 1.	736	704	691	677	651	653	645	641	31
CRmg < 1.	662	664	671	672	662	649	639	636	13
Ratios from Break Points									
CRbpe < .31	370	295	275	243	258	250	226	232	44
CRbpg < .58	395	298	287	251	255	245	233	220	52
CRbpe < .52	**504**	**462**	**494**	**488**	**506**	**505**	**504**	**505**	**14**
CRbpg < .77	**517**	**504**	**493**	**525**	**499**	**511**	**496**	**495**	**11**
CRbpe < 1.	672	702	735	765	732	744	749	752	28
CRbpg < 1.	669	685	740	770	733	742	758	757	34
G & W's Results									
CR < 10%	585	470	404	321	259	216	148	127	150
G < 33rd percentile	367	356	326	328	324	328	325	340	15

Notes: me = means from eigenvalue (CI) routine, mg= means from G method
bp = break point

As also shown at the bottom of Table 5, Golden and Wang in their study produced about the same discriminatory power (standard deviation of 15) for their 33rd percentile of G and much worse performance for the 10% rule (standard deviation and 150). What we do not know from their study is the degree of perturbation they used across matrix size. Assuming no truncation to 1/9 or 9, we can calculate that the theoretical average perturbation for the n(n-1)/2-(n-1) comparisons they changed will be about .66, 1., 1.2, 1.3, 1.42, 1.5, 1.55 and 1.6 from n=3 to n=10 respectively. Exacerbating this fact is the larger variance which occurs at smaller matrix sizes (see Table 3). Thus for small n, their consistency rule will tend to reject more matrices and require quite accurate comparisons for acceptance. When n is larger, their procedure quickly stabilizes to the same approximate level of comparison accuracy in order for acceptance to occur.

In this study, where each matrix size has the same the degree of perturbation in the validation sample (1.25), the cutoffs using mean values (CRm) have declining acceptance rates across matrix sizes. Only CRmg < 1. has consistent results (standard deviation of 13). but it not totally suitable, because it has an acceptance rate much higher than the 10 percent rule. On the other hand, the cut-off ratios based upon the break points produce very good results. Both CRbpe <.52 and CRbpg < .77 (the break point between

Table 6 -- Average Perturbation Levels for Accepted Matrices

Measure of Effectiveness	Matrix Size and Average Perturbation Size								Std Dev
	3 ±1.24	4 ±1.24	5 ±1.24	6 ±1.24	7 ±1.24	8 ±1.24	9 ±1.25	10 ±1.24	
CR < 10%	1.06	.97	.88	.81	.76	.75	.74	.74	.11
G < 33rd percentile	.96	.88	.83	.80	.74	.72	.70	.70	.09
Modified CR < 10%	.88	.86	.88	.81	.76	.75	.74	.74	.11
Ratios from Means									
CRme < .31	.97	.81	.69	.60	.57	.55	.51	.49	.16
CRmg < .55	.96	.77	.65	.57	.58	.56	.49	.48	.15
CRme < .47	**1.03**	**.89**	**.78**	**.71**	**.71**	**.70**	**.67**	**.69**	**.12**
CRmg < .73	**.98**	**.90**	**.82**	**.76**	**.70**	**.70**	**.67**	**.66**	**.11**
CRme < 1.	1.08	1.04	.98	.96	.92	.92	.91	.90	.06
CRmg < 1.	1.05	1.04	.99	.98	.94	.93	.92	.91	.05
Ratios from Break Points									
CRbpe < .31	.94	.81	.72	.65	.64	.62	.58	.58	.12
CRbpg < .58	.98	.83	.74	.67	.64	.61	.58	.56	.13
CRbpe < .52	**.99**	**.90**	**.85**	**.80**	**.77**	**.76**	**.75**	**.75**	**.08**
CRbpg < .77	**1.02**	**.95**	**.87**	**.88**	**.80**	**.79**	**.77**	**.76**	**.09**
CRbpe < 1.	1.05	1.04	1.01	1.05	1.00	1.00	1.00	1.00	.02
CRbpg < 1.	1.06	1.05	1.04	1.06	1.01	1.02	1.03	1.03	.01

perturbation k±1 vs. k±2) have the lowest standard deviations of their respective measures. Moreover, the acceptance rates at this level are very close to the number accepted by the 10 percent rule at n>5 where the 10 percent rule is stable.

Another way to judge the performance of each rule is to look at the average perturbation level of matrices judged to be acceptable (Table 6). Again, we can see that the greatest stability across matrix size (lowest standard deviations) is produced by the cut-off ratios from break points. What is somewhat disheartening about all methods is that the lower cutoff levels accept small-sized matrices with average perturbation higher than what they are supposed to discriminate against. This can be attributed to the high positive skewness and closeness of perturbation means at small matrix sizes. What is encouraging, however, is that for the highest cut-off level (the break point between perturbation k±2 vs. k±3), there is sufficient distance from the distributions to allow almost perfect discrimination. At that level, the number of accepted matrices is close to the 250 validation matrices from each of the first three perturbation levels and the average perturbation of the accepted matrices is almost identical to the average of those three distributions (i. e. equal to 1.)

DISCUSSION

The simulation procedure used in this study controlled the level of departure from pure consistency for both developmental and validation samples. It has shown that stable discrimination for all matrix sizes can be achieved if we use a consistency ratio based upon the breakpoint between distributions which have an average comparison perturbation of 1.5 and 2. Using these breakpoints as the upper limit for tolerable consistency, we can create a new type of ratio in which unity represents the dividing line between acceptable and unacceptable comparisons. This new demarcation line is more stable across matrix sizes than the 10 percent rule which is based upon departures from pure randomness. Moreover, the breakpoints between less severely perturbed distributions can be used by AHP software designers to indicate additional gradations of more satisfactory levels of consistency.

Although the new measure works well for both CI and G, there is a slight advantage of using G as the measure of inconsistency. Not only is it easy and quick to calculate, it also results in more uniform distributions when matrices are perturbed. But since these benefits are marginal, it is unlikely that programmers will make the change.

If the G measure is adopted, then it is recommended that it be used as a breakpoint consistency ratio. Golden and Wang's 33 percentile rule gives relatively good discrimination, but it is also based upon unknown and potentially variable perturbation levels for different matrix sizes. The consistency ratio rules developed in this study are based upon known levels of perturbation for each individual comparison.

Except for the first level of perturbation, the size of the disturbances from pure consistency were values from uniform integer distributions. An alternate procedure would be to take each perturbation of the consistent comparison from a normal or other symmetrical distribution. If non-integer comparisons are allowed in the AHP procedure, then selection from normal distributions may better represent the finer gradations of departure from consistency. In this study, the consistent comparisons were non-integer to start with and the integer perturbations maintained the non-integer value.

Although we have chosen an average perturbation of between 1.5 and 2 as the breakpoint for acceptable consistency, we have very little knowledge of how people actually behave when using AHP. Tests of AHP have determined how close users come to replicating a known priority vector. If we knew in such experiments the true priority, then we should also know the true paired comparison. Using experiments with known stimuli under different dispersions and matrix sizes, we should determine how close people come to emulating these true paired comparisons. That way, we will be able to specify average perturbation levels that people achieve in practice. We could then use that information to chose standards for acceptable consistency ratios.

Finally, we must consider this new consistency ratio in light of the entire hierarchy. In terms of calculating the consistency ratio for the hierarchy, we would directly substitute the new measure into the overall calculations. The only difference is that we would use unity rather than 10 percent to determine whether sufficient consistency has been achieved. For situations where we generate priorities from incomplete comparisons, we could use prediction equations to determine CI or G. Except for a different interpretation of the consistency ratio, there are few problems to introduce the new measure.

CONCLUSIONS

This study has approached the determination of the consistency ratio as a relative departure from pure consistency. The traditional measure which looks upon the ratio as a sufficient departure from randomness provided variable discrimination at different matrix sizes. By placing the upper limit of tolerable consistency in the denominator, we develop a ratio where unity is the dividing point between acceptable and unacceptable consistency. Other demarcations less than unity represent better levels of consistency.

Such a new consistency ratio works well for both CI and G as the measures of consistency. It is also easy to implement. Although behavioral research and subsequent experiments may improve upon how this relative departure from consistency is used, there is sufficient evidence from this study to suggest its adoption at the present time.

REFERENCES

Bruce L. Golden and Qiwen Wang, "An Alternative Measure of Consistency," Golden, Bruce L., Wasil, Edward A. and Harker, Patrick T. (eds.) The Analytic Hierarchy Process: Applications and Studies, Heidelberg, Springer-Verlag, 1989, pp. 68-81..

Eric F. Lane and William A Verdini, "A Consistency Test for AHP Decision Makers," Decision Sciences, Vol. 20, 1989, pp. 575-590.

Saaty, Thomas L. (1977) "A Scaling Method for Priorities in Hierarchical Structures," Journal of Mathematical Psychology, Vol. 15, No. 3, pp. 234-81.

__________, (1980) The Analytic Hierarchy Process, New York, McGraw Hill