A Comparison of the Analytic Hierarchy Process and the Geometric Mean Procedure for Ratio Scaling

The development of analytical procedures and experimental techniques for constructing ratio measurement scales has long been a major topic and important challenge in psychophysics and other areas of psychology. The interest in ratio scales is obviously related to their high level of invariance (unique up to a similarity transformation) and the associated statistical operations they allow one to perform (Stevens, 1946). To obtain ratio scales, Stevens and his colleagues (Stevens, Mack, & Stevens, 1960) developed the "cross modality matching" paradigm, which was originally applied to variables with a corresponding physical continuum and later generalized to social and other psychological stimuli (e.g., Lodge, 1981). Although Stevens' techniques have been widely used in numerous areas (e.g., Stevens, 1972, 1975; Lodge, 1981), their precise characterization is still a topic of debate among measurement specialists (e.g., Shepard, 1981).

Following two decades of development of ordinal level scales, which culminated in the development of nonmetric multidimensional scaling, there has recently been a renewed interest in the problem of ratio scale measurement. To a large extent, the interest in this problem is due to the development of a new ratio scaling procedure and its successful use in a variety of experimental and practical situations. The most salient characteristic of this method was apparently first discovered by Gulliksen (1959), but the procedure was fully developed, investigated, described, and applied by Saaty (1977, 1980).

Saaty's Analytic Hierarchy Process

Consider a set of n stimuli with unknown scale values, $\underline{s}^1 = (s_1,\dots,s_n)$. Following a process of pairwise comparisons, an n x n matrix of ratios, R, is constructed such that

(1) $$r_{i,j} = s_i/s_j, \quad i,j=1,\dots n,$$

All the entries in R are positive, satisfying the reciprocity condition: $r_{j,i} = 1/r_{j,i}$. In reality, this constraint is artificially enforced because typically only $n(n-1)/2$ judgments are obtained and the remaining ratios are calculated by the reciprocal transformation and by assuming that $r_{i,i} = 1$ $(i=1,\dots,n)$.

Saaty's procedure is based on a relatively simple and elementary property of R: when postmultiplied by the vector of scale values, the result is a vector related to the scale values by a constant. This constant turns out to be the size of the matrix R. Thus,

(2) $$R\underline{s} = n\underline{s}.$$

This formulation suggests that the unknown scale values can be obtained by an eigenvalue-eigenvector decomposition of R. Saaty's solution to this problem when the data are perturbed is to use the normalized right eigenvector associated with the largest eigenvalue of R, denoted by λ_{max}, as an estimate of the scale values.

The matrix R is said to be <u>consistent</u> if the ratios $r_{i,j}$ satisfy

(3) $$r_{i,j} = r_{i,k}r_{k,j} \qquad i,j,k=1,\ldots n$$

for all i, j, and k. A consistent matrix is of unit rank. Moreover, its only nonzero eigenvalue must be n (see Eq. 2). It can be shown that $\lambda_{max} > n$ when consistency is violated. Therefore, Saaty recommends using the normalized difference

(4) $$\mu = (\lambda_{max}-n)/(n-1)$$

as a measure of inconsistency. If R is consistent, $\mu = 0$; if not, μ is monotonically increasing in the magnitude of departure from consistency (for more details of the properties of this method see Saaty, 1977, 1980).

Saaty's analytic hierarchy process (AHP) has attracted much attention, having been successfully applied in such diverse areas as marketing (Wind & Saaty, 1980), political science (Saaty & Bennett, 1977), and the measurement of subjective probabilities (Yager, 1979). However, the AHP has had its share of criticism. In particular, Johnson, Beine, and Wang (1979) have pointed out that for inconsistent matrices of order $n \geq 4$ the solution is not invariant under transposition. In other words, the right eigenvector of R' (i.e., the left eigenvector of R) is not necessarily the reciprocal of the right eigenvector of R. This may cause difficulties in the interpretation of the scale values (see also Budescu, 1984).

<u>The Geometric Means Procedure</u>

Another criticism was leveled by Williams and Crawford (1980), who argued that, unlike most estimation procedures, Saaty's procedure does not optimize a well defined loss function. Moreover, it requires complex calculations as the maximal eigenvalue and the associated eigenvector are calculated by an iterative technique. Therefore, Williams and Crawford proposed an alternative procedure, called the geometric means procedure, which is invariant under transposition and can easily be calculated by hand.

It has long been recognized that the geometric means (GMs) are least squares estimates of the logarithms of the scale values (e.g., Torgerson, 1958). Williams and Crawford (1980) have shown that if the true scores are perturbed by independent lognormally distributed errors with zero mean and variance σ^2, the GMs are also maximum likelihood estimates of the scale values (for a theoretical justification of the lognormal distribution of errors in the judgment of dissimilarities see Ramsay, 1977). To measure the inconsistency of R, Williams and Crawford proposed using the residual mean square

$$(5) \qquad S^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} [\ell n(r_{i,j}) - \ell n(GM_i/GM_j)]^2/(n-1)(n-2),$$

where S^2 is an estimate of σ^2. As in Eq. 4, $S^2 = 0$ if R is consistent. As the departure from consistency increases, S^2 increases monotonically.

A Comparison Based on Perturbed Scale Values

How do the two scaling procedures compare? For a consistent judgment matrix of any order or for any reciprocal matrix of order $n \leq 3$, the eigenvector of R (associated with λ_{max}) and the vector of row GMs are equal except for a similarity transformation. The procedures differ, however, when consistency is violated. To compare the two procedures, Williams and Crawford conducted a Monte Carlo study in which reciprocal matrices of order 5, 7, and 10 were perturbed by multiplying each ratio S_i/S_j by an error e_{ij}. Two types of errors were investigated. In the first, the errors e_{ij} were drawn from a lognormal distribution with zero mean and variance σ^2. In the second, the errors were drawn from a population of ratios of uniform random variables with the same mean and variances as in the first case. Five values of σ^2 were examined, and each condition was repeated 1000 times. In all cases the GM procedure outperformed Saaty's procedure according to both the least squares and the log least squares criteria. Moreover, the relative advantage of the GM procedure increased with both the size of the judgment matrix R and the variability of error. Table 1, which is taken from Tables 2 and 3 of Williams and Crawford, shows the percentage of cases (matrices) in which the GM procedure outperformed the AHP.

In the "almost consistent case" (top row of Table 1), when the variability of error is very small, the two solutions are equally good. But as the variability of error and the size of the response matrix increase, the GM's advantage

Table 1

Percentage of Cases in which the GM Procedure Provides Better Least Squares Estimates of Scale Values than the AHP

	Lognormal Errors				Ratio of Uniform Errors			
		n				n		
σ^2	5	7	10	Total	5	7	10	Total
.01	52	51	50	51.00	51	49	50	50.00
.16	54	56	57	55.67	51	52	53	52.00
.64	50	54	57	53.67	57	62	66	61.67
1.00	56	62	70	62.67	59	68	73	66.67
1.69	58	66	71	65.00	71	81	92	81.33
Total	54.00	57.80	61.00	57.60	57.80	62.40	66.80	62.33

Note: Each cell is based on 1000 cases.

over Saaty's procedure is enhanced. Table 1 suggests that nothing is lost by using the GM procedure in cases where Saaty's AHP works well. There seems to be a real potential gain in employing the GM procedure in other more erratic cases.

A particularly attractive feature of Saaty's AHP is the availability of a consistency index μ (Eq. 4) which differentiates between judgment matrices that can be maintained and interpreted, and judgment matrices that must be rejected as "randomly generated data." As a rule of thumb, Saaty (1977) proposed to fix μ at 0.1. Later (1980, 1983), he recommended using a consistency ratio denoted by C.R. and defined by

(6) $$\text{C.R.} = \mu/\mu_n,$$

where μ_n is an empirical measure computed from Eq. (4) for 500 randomly generated reciprocal matrices of order n. The new rule of thumb (presumably for 1-9 response scales) is to accept only judgment matrices for which C.R. $\leq$ 0.2.

Although Williams and Crawford have suggested the statistic S^2 as a measure of inconsistency for the GM procedure, no corresponding rules are available for its use. This, perhaps, reflects the shortage of applications of the GM procedure. More importantly, because μ and S^2 are based on two different models of scaling matrices of ratio judgments, and do not necessarily reflect the same properties of the data, it is still an open question how the two procedures compare in the null case--when the data are known to be random (rather than consistent judgments perturbed by error).

Objectives

A major purpose of the present note is to contrast and evaluate the two scaling procedures in the null case and provide guidelines for their use with real data. In this case both scaling procedures are expected to reject by means of μ and S^2 the null hypothesis of consistency. The methodology employed in the present paper is very similar to that used by Saaty (1977, 1980): a large number of reciprocal matrices consisting of random entries are generated, solutions and indices of inconsistency are calculated, and the rejection rules are compared to each other.

A second purpose of this note is to generalize the findings from the comparison of the two scaling procedures to a nonnumerical method of obtaining ratio judgments. Saaty's procedure obtains directly numerical estimates of ratios in the tradition of Stevens' magnitude estimation. Furthermore, Saaty strongly recommends restricting the response scale to the positive integers 1 through 9 and their reciprocals (a total of 17 possible different values). Another equally popular experimental procedure for eliciting ratio judgments is the "constant sum" method (e.g., Torgerson, 1958) in which a judge is required to divide a constant number of units between two stimuli in accordance with their ratio. For example, a 90:10 allocation of 100 units reflects a ratio of 90/10 = 9, and a 50:50 division gives rise to a ratio of 1. In certain applications, such as encoding subjective probabilities, it is preferable to avoid using numerical responses because of their inherent bias. Instead, graphical methods such as placing a sliding marker along a bounded straight line or adjusting a two-color probability wheel are recommended (e.g., Spetzler & Stael von Holstein, 1975; Wallsten & Budescu, 1983). Then the ratio judgments can be derived directly from the relative lengths (areas) of the two segments of the line (wheel) as in the direct estimation method.

When graphical techniques are employed, the derived ratio judgments are not necessarily identical to those obtained from direct ratio judgments. In particular, they are not necessarily integer values (there is no way to divide 100 units to obtain a 5:1 ratio), and they are likely to yield extreme ratios near the end points (e.g., 100 units may yield ratios as high as 99 and 49) and more densely clustered ratios elsewhere. Because the scale values and, consequently, the two inconsistency measures are sensitive to the response scale (see Saaty, 1980), we plan to compare the two scaling procedures for the "graphical constant sum" method as well.

Method

A Monte Carlo study was conducted to compare the AHP and GM procedures for scaling ratio judgments. Thirty different conditions were generated by factorially combining three independent variables:

1. The number of stimuli to be scaled: $n = 4,6,8,10,12$.
2. The experimental method for eliciting ratio judgments: direct estimation and constant sum.
3. The number of different responses allowed: $k = 17,25,99$.

The third factor of the present design reflects an assumption regarding the level of differentiation and precision that the judge may achieve when selecting a particular response. For the direct estimation method the three values of k imply that the integers 1-9 (as suggested by Saaty), 1-13, 1-50 and their reciprocals are used, respectively. For example, if $k = 17$ and the (i,j) cell of R is considered, an integer x is chosen randomly from the integers 1,...,9 such that $r_{j,i} = x$ and $r_{ij} = 1/x$.

When the constant sum method is employed, it is assumed that the judge can divide the total number of units specified by the experimenter (or the total length of a line) in only a finte number (k) of equally spaced categories. Each response is assumed to be the mid-point of an interval of unit length. The function relating the numerically calculated ratios r_{ij} to the nonnumerical responses is convex. For example, suppose a line is divided into k=17 equally spaced intervals. If the judge places the marker on the 9th interval, $r_{ij} = 8.5/8.5 = 1$. A marker placed on the 11th interval yields a ratio of $10.5/6.5 = 1.615$, and one placed on the right-hand (17th) interval yields the highest possible ratio of $16.5/0.5 = 33$. The first two levels of the third factor (k=17, 25) are representative of the constrained response scales used in the psychological literature. The third level (k=99) is included as a reasonable approximation to the unconstrained situation in which the judge may select any real number in making his or her judgment.

For each of the 30 conditions in the 3-way factorial design described above, 1000 matrices were generated by independently choosing $n(n-1)/2$ uniformly distributed integers within the range of values dictated by k and the method for eliciting ratio judgments. The numbers were randomly placed in the cells of an n x n matrix (but excluding the diagonal entries, which are all 1's), and their reciprocal values were assigned to the corresponding transposed positions. For each matrix R_q (q=1,...,1000) in each of the 30 conditions four solutions were obtained:

1. The geometric means of R_q's rows: $\underline{s}_{gq}$,
2. The right eigenvector of R_q: $\underline{s}_{rq}$,
3. The left eigenvector of R_q: $\underline{s}_{\ell q}$,
4. The geometric mean of the right and left eigenvectors of R_q: $\underline{s}_{mq}$.

Only the first and fourth solutions are invariant under transposition of R_q. The estimated scale values under each model and the two measures of inconsistency μ and S^2 were then computed and printed.

Results

Table 2 presents a summary of several characteristics of the sampling distributions of μ and S^2 for the direct estimation method. For both statistics, each row of Table 2 displays two measures of central tendency (mean and median), a measure of variability (standard deviation), and a measure of skewness defined as (e.g., Nichols & Gibbons, 1979):

$$\varepsilon = 3\ (\text{Mean} - \text{Median})\ /\ \text{standard deviation}.$$

ε takes on values between -3 and 3; it is centered around 0 when the distribution is symmetric.

Table 3 shows a similar summary for the constant sum method of eliciting ratio judgments.

Both Tables 2 and 3 show that the two location parameters for μ increase in both k and n, whereas for S^2 they only increase in k. For a given k, and regardless of which method is used to elicit responses, μ increases on the average as the number of stimuli grows, whereas S^2 remains unchanged. For both μ and S^2 and for both methods, the standard deviations decrease in n and, with some exceptions in Table 2, increase in k. Inspection of both the variability and skewness measures shows that all 60 sampling distributions become more stable (i.e., less variable) and symmetric as the number of stimuli increases. A comparison of Tables 2 and 3 demonstrates that both effects are stronger for the direct estimation case.

Table 2

Summary Statistics of the Sampling Distributions

of μ and S^2 for the Direct Estimation Method

		μ				S^2			
k	n	Mean	Median	S.D.	Skewness	Mean	Median	S.D.	Skewness
	4	0.895	0.653	0.824	0.881	2.575	2.176	1.886	0.635
	6	1.458	1.316	0.774	0.550	2.718	2.638	1.071	0.224
17	8	1.615	1.548	0.583	0.345	2.602	2.566	0.702	0.154
	10	1.764	1.740	0.464	0.155	2.628	2.601	0.533	0.152
	12	1.823	1.817	0.377	0.048	2.585	2.578	0.412	0.051
	4	1.264	1.084	0.925	0.584	3.472	3.223	2.156	0.347
	6	1.840	1.870	0.630	-0.143	3.477	3.527	1.044	-0.144
25	8	2.090	2.092	0.448	-0.013	3.527	3.538	0.716	-0.046
	10	2.231	2.236	0.318	-0.047	3.544	3.537	0.490	0.043
	12	2.308	2.314	0.263	-0.068	3.550	3.568	0.390	-0.138
	4	5.473	4.329	4.513	0.760	9.265	8.505	5.753	0.396
	6	8.680	8.883	2.877	-0.212	9.537	9.703	2.760	-0.180
99	8	9.649	9.735	1.799	-0.143	9.487	9.510	1.648	-0.042
	10	10.269	10.273	1.293	-0.009	9.606	9.619	1.174	-0.033
	12	10.541	10.556	1.039	-0.043	9.540	9.575	0.911	-0.155

Table 3

Summary Statistics of the Sampling Distributions of μ and S^2 for the Constant Sum Method

		μ				S^2			
k	n	Mean	Median	S.D.	Skewness	Mean	Median	S.D.	Skewness
	4	1.110	0.760	1.199	0.876	2.996	2.440	2.319	0.719
	6	1.704	1.418	1.127	0.775	2.929	2.756	1.291	0.402
17	8	1.939	1.805	0.846	0.475	2.842	2.774	0.820	0.249
	10	2.167	2.055	0.736	0.457	2.868	2.842	0.650	0.120
	12	2.291	2.228	0.608	0.311	2.859	2.837	0.506	0.130
	4	1.159	0.743	1.309	0.953	3.056	2.395	2.456	0.807
	6	1.768	1.486	1.194	0.709	3.003	2.829	1.336	0.391
25	8	2.140	1.966	1.025	0.509	2.975	2.872	0.886	0.349
	10	2.140	1.966	1.025	0.509	2.975	2.872	0.886	0.349
	12	2.582	2.488	0.791	0.357	2.995	2.967	0.566	0.148
	4	1.306	0.787	1.636	0.952	3.286	2.518	2.847	0.809
	6	2.115	1.563	2.122	0.780	3.228	2.931	1.637	0.544
99	8	2.621	2.139	1.815	0.797	3.195	3.023	1.113	0.464
	10	2.986	2.633	1.562	0.678	3.224	3.166	0.853	0.204
	12	3.300	3.035	1.431	0.556	3.208	3.168	0.662	0.181

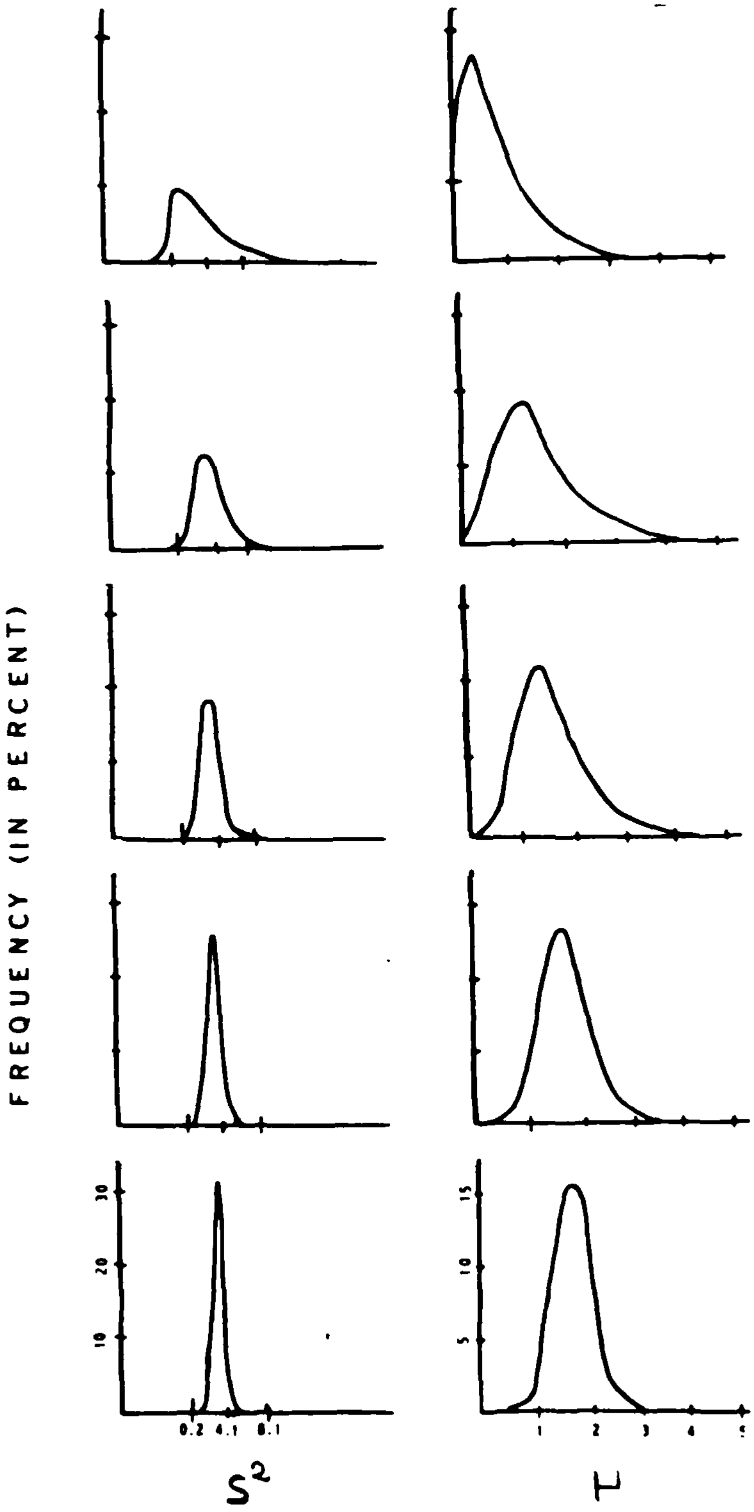

Figure 1: Sampling distributions of μ and S^2 under direct Estimation for k = 17, and n = 4, 6, 8, 10, 12.

To illustrate the increasing stability of the sampling distributions as the number of stimuli increases, Figure 1 portrays 10 sampling distributions under the direct estimation for $k=17$ (the response scale advocated by Saaty) and $n = 4, 6, 8, 10$, and 12. The right panel of the figure displays the sampling distributions for the AHP and the left panel for the GM procedure. The reduction in variability and skewness is evident in the figure. Tables 4 and 5 present the critical values of the null sampling distributions (at $\alpha = 0.10$, 0.05, and 0.01) for testing the null hypothesis of randomly generated responses. Because both μ and S^2 vanish when the ratio judgments are consistent, the null hypothesis is rejected (and the derived scale values may be safely accepted and properly interpreted) at a given level whenever a value smaller than that listed in the appropriate table is obtained.

Because Tables 4 and 5 are most important for practical purposes, it is desirable to generalize their results to values of n and k other than those examined in the present study. We have achieved this purpose by fitting relatively simple multiple regression equations based on n, k, or some monotonic transformations of these two parameters. For the constant sum method, the critical values of the sampling distribution of μ are approximated by

$$F_{.10} = -1.2043 + 0.0979n + 0.6719 \log(n) + 0.0008k,$$
$$F_{.05} = -0.9728 + 0.1150n + 0.4243 \log(n) + 0.0003k,$$
$$F_{.01} = -0.4143 + 0.1715n \mp 0.1815 \log(n),$$

whereas the critical values for the sampling distribution of S^2 are given by:

$$F_{.10} = -2.7024 - 0.1923n + 2.9253 \log(n) + 0.0006k,$$
$$F_{.05} = -2.6871 - 0.1351n + 2.5866 \log(n) + 0.0003k,$$
$$F_{.01} = -1.8826 + 0.0411n + 1.3067 \log(n) + 0.0006k.$$

The six equations above fit the Monte Carlo results extremely well; they all have adjusted squared multiple correlations of at least 0.99, and all residuals are less than 0.085.

For the direct estimation procedures, the critical values of μ are approximated by:

$$F_{.10} = -3.1870 + 0.9711n - 0.0325k - 0.0605n^2 + 0.0109nk$$
$$F_{.05} = -2.1582 + 0.6270n - 0.0355k - 0.0429n^2 + 0.0144nk,$$
$$F_{.01} = -0.4382 + 0.1744n - 0.0511k - 0.0155n^2 + 0.0116nk.$$

And the critical values for S^2 are approximated by

$$F_{.10} = -4.5034 + 1.0793n + 0.0688k - 0.0620n^2 + 0.0062nk - 0.0005k^2,$$
$$F_{.05} = -4.2697 + 1.0044n + 0.0544k - 0.0567n^2 + 0.0065nk - 0.0005k^2,$$
$$F_{.01} = -3.3187 + 0.7788n + 0.0190k - 0.0434n^2 + 0.0077nk - 0.0003k^2.$$

Table 4

Critical Values of the Sampling Distributions of μ and S^2 for the Direct Estimation Method

			μ			S^2	
k	n	.10	.05	.01	.10	.05	.01
	4	0.148	0.090	0.035	0.560	0.346	0.136
	6	0.596	0.471	0.237	1.403	1.177	0.756
17	8	0.903	0.804	0.581	1.701	1.577	1.220
	10	1.174	1.037	0.847	1.952	1.783	1.482
	12	1.324	1.211	1.048	2.039	1.940	1.721
	4	0.246	0.153	0.048	0.896	0.577	0.189
	6	0.944	0.773	0.414	2.055	1.749	1.065
25	8	1.488	1.348	1.044	2.601	2.353	1.869
	10	1.832	1.715	1.347	2.916	2.750	2.313
	12	1.962	1.881	1.676	3.071	2.890	2.560
	4	0.769	0.569	0.204	2.535	1.926	0.744
	6	4.841	3.039	1.346	5.716	4.762	2.783
99	8	7.298	6.528	5.068	7.328	6.543	5.433
	10	8.634	8.134	7.118	8.054	7.583	6.860
	12	9.171	8.755	8.057	8.365	7.934	7.155

Table 5

Critical Values of the Sampling Distributions of μ and S^2 for the Constant Sum Method

		μ			S^2		
k	n	.10	.05	.01	.10	.05	.01
	4	0.164	0.103	0.036	0.610	0.393	0.142
	6	0.604	0.487	0.276	1.410	1.167	0.675
17	8	0.999	0.835	0.534	1.856	1.604	1.118
	10	1.301	1.56	0.944	2.036	1.889	1.632
	12	1.553	1.409	1.188	2.210	2.094	1.851
	4	0.160	0.095	0.032	0.593	0.363	0.125
	6	0.620	0.479	0.238	1.455	1.170	0.655
25	8	1.033	0.874	0.604	1.891	1.641	1.224
	10	1.367	1.184	0.887	2.151	1.933	1.563
	12	1.671	1.445	1.201	2.288	2.101	1.821
	4	0.166	0.094	0.034	0.620	0.359	0.134
	6	0.601	0.455	0.270	1.410	1.144	0.698
99	8	1.032	0.830	0.609	1.871	1.594	1.245
	10	1.456	1.237	0.952	2.173	2.002	1.646
	12	1.784	1.534	1.122	2.396	2.202	1.863

Table 6

Measures of Agreement Between μ and S^2

		Direct Estimation				Constant Sum			
k	n	τ	$P_{.10}$	$P_{.05}$	$P_{.01}$	τ	$P_{.10}$	$P_{.05}$	$P_{.01}$
	4	.972	99	100	100	.971	99	100	80
	6	.876	92	92	100	.890	95	94	90
17	8	.819	85	76	80	.812	90	84	100
	10	.797	81	80	90	.763	75	74	80
	12	.794	80	78	60	.702	77	80	70
	4	.953	99	100	100	.969	99	100	100
	6	.841	90	90	90	.874	92	92	90
25	8	.835	83	72	80	.809	90	86	90
	10	.807	80	76	90	.784	75	82	90
	12	.820	81	82	70	.763	63	60	80
	4	.925	97	96	100	.971	100	98	90
	6	.736	87	86	90	.881	93	92	100
99	8	.704	69	72	50	.806	85	84	80
	10	.707	67	58	80	.755	83	82	80
	12	.714	70	66	20	.728	72	76	80

The fit of these six equations is not as impressive as the fit reported above for the constant sum method. All the adjusted squared multiple correlations exceed 0.97, but some of the residuals for the case k=99 seem larger than desired.

The agreement between μ and S^2 is summarized by several measures in Table 6. The two columns labelled τ, one for each method, report the values of Kendall's rank correlation between the 1000 pairs of μ and S^2 calculated separately for each of the 30 conditions. It is well known (Marascuilo & McSweeny, 1977) that τ can be provided a probabilistic interpretation: it is the difference between the probabilities of finding a concordant and a discordant pairs of values under random sampling. The other six columns focus on the regions of rejection defined in Tables 4 and 5. Specifically, they present the percentage of cases for which the null hypothesis of randomly generated judgments is rejected by S^2 <u>and</u> μ for α = 0.10, 0.05, and 0.01.

It can safely be concluded from Table 6 that the agreement between μ and S^2 is quite high; the mean rank order correlation between μ and S^2 is 0.813 for the direct estimation and 0.832 for the constant sum method, and the corresponding mean percentage of agreement at α = 0.05 is 81 and 86, respectively. Table 6 shows that the two measures of inconsistency are in closer agreement in the constant sum method. This slight advantage of the constant sum over the direct estimation method is attributed to the results obtained for k=99. Table 6 further shows a strong effect of the number of stimuli--as n increases the agreement between μ and S^2 declines.

Following our examination of the agreement between μ and S^2, we turn next to investigate the relations between the four solutions to the scaling problem: $\underline{S}_g$, $\underline{S}_\ell$, $\underline{S}_r$, and $\underline{S}_m$.

Tables 7 and 8 present the median Pearson product moment correlations between any two solutions across the 1000 replications. Table 7 shows the median correlations for the direct estimation method, whereas the corresponding results for the constant sum method are presented in Table 8. All the correlations are high, indicating that the four solutions yield similar scale values in the null case as well as in the unidimensional perturbed case (Williams & Crawford, 1980). An examination of the two tables reveals several effects. Here again, as in Table 6, there is a strong and consistent inverse relationship between the number of stimuli (n) and the degree of association between the various pairs of solutions. Also similar to Table 6, there are

Table 7

Median Correlations* Between the Four Solutions for the Direct Estimation Method

k	n	$S_g x S_\ell$	$S_g x S_m$	$S_g x S_r$	$S_\ell x S_m$	$S_\ell x S_r$	$S_m x S_r$
	4	987	997	989	994	979	996
	6	918	970	914	955	791	955
17	8	864	951	857	923	678	921
	10	828	931	829	911	592	891
	12	795	923	811	893	517	863
	4	991	996	990	997	988	998
	6	922	969	919	965	805	945
25	8	893	962	897	946	733	920
	10	882	961	883	933	686	906
	12	888	962	880	935	691	906
	4	994	996	987	999	991	997
	6	884	939	858	958	734	922
99	8	849	917	825	939	674	896
	10	835	913	821	928	632	890
	12	846	922	818	924	635	887

*Decimal points are omitted.

no systematic differences among the correlations due to the response scale (k). A comparison of Tables 7 and 8 shows that under the direct estimation method the correlations between the solutions are slightly higher than under the constant sum method.

When the six correlations within a condition (row) are compared to one another in either of the two tables, an interesting pattern emerges. In all the 30 cases examined the lowest correlation is between the left and right eigenvectors ($\underline{S}_\ell * \underline{S}_r$), and, with the exception of k=99 in Table 7, the highest correlation is between the geometric mean solution and the geometric mean of the two eigenvector solutions ($\underline{S}_g * \underline{S}_m$). It may be recalled that of the four solutions, only $\underline{S}_g$ and $\underline{S}_m$ are invariant under transposition of the judgment matrix R.

Table 8

Median Correlations* Between the Four Solutions for the Constant Sum Method

k	n	$S_g x S_\ell$	$S_g x S_m$	$S_g x S_r$	$S_\ell x S_m$	$S_\ell x S_r$	$S_m x S_r$
	4	987	996	988	993	977	996
	6	917	970	913	945	789	958
17	8	852	940	834	896	627	922
	10	814	922	791	881	538	892
	12	765	895	754	864	479	867
	4	986	996	986	994	975	995
	6	914	972	915	943	789	962
25	8	855	940	817	844	605	920
	10	808	909	796	869	535	899
	12	768	832	735	841	431	868
	4	926	976	934	945	845	995
	6	918	970	912	941	788	960
99	8	834	931	822	876	586	927
	10	791	896	745	824	479	903
	12	751	852	663	791	869	876

*Decimal points are omitted.

Discussion

The results of the present study must be discussed from two perspectives, the practical and the methodological. At the practical level, the results provide a much needed service to those who employ ratio scaling procedures routinely in their work and who are often concerned with the reliability and consistency of their data. The results presented in Tables 2 through 5 provide these researchers a sound basis for detecting inconsistent judgments and subjects. They supply decision rules for accepting or rejecting judgment matrices for both the AHP and GM procedures. In the former case, we contend that use of the traditional approach to hypothesis testing, in which the null hypothesis is rejected with a predetermined probability of type I error, is superior to any of the rules of thumb advocated by Saaty. In the latter case, our results fill a gap in the long line of psychometric studies on the properties of the geometric means.

Like any other simulation, our results are limited to those combinations of parameters tested. As such, they lack the generality achievable by a theoretical development. However, the values of n and k employed in the present study are representative of the choices usually encountered in applications, so that many users should find our tables useful. Those using different response scales and a larger number of stimuli will benefit from the approximation formulas developed for the critical points of the sampling distributions. Undoubtedly, these formulas may be further refined in similar studies covering a larger range of parameter values.

We point out to the more conservative user that the results in Tables 2 and 3 can be used to approximate the mean and variance of the sampling distributions which, in conjunction with Tchebycheff's inequality, could be used to determine more conservative decision rules. For example, in the direct estimation method the mean and standard deviation of S^2 can be approximated by

$$\bar{S}^2 = 1.2259 + 0.0129n + 0.0825k,$$

$$\sigma(S^2) = [2.440 - 0.335n + 0.006k + 0.016n - 0.001nk] .$$

Both equations fit the data well ($R_{adj.} > 0.975$). Similar approximations can be generated for other cases.

At a more general methodological level, we conclude that there is a very high level of agreement between the AHP and GM procedures in the null case. We base this conclusion on high correlations between the estimated scale values (Tables 7 and 8), the large concordance between the rank orderings of the various matrices by means of μ and S^2, and the high agreement between the two rejection rules for various levels of α (Table 6). It was pointed out in the introduction that one interpretational problem of the AHP stems from its lack of invariance under transposition of R. The magnitude of this problem is well documented in Tables 7 and 8, where the right and left eigenvectors are shown to intercorrelate lower than any other pair of solutions. This finding reinforces the warning of Johnson, Beine, and Wang (1979) regarding the use of the right eigenvector. It is reassuring to note that in the same two tables the highest correlations are typically those between the two invariant solutions.

When our results are considered in conjunction with the findings of Williams and Crawford (1980) and the objection raised by Johnson, Beine, and Wang (1979), a strong case can be made in favor of the GM as a better and more convenient procedure than the AHP for scaling pairwise ratio judgments. In the null case as well as in the consistent case, the two procedures are practically indistinguishable. But the geometric means are easier to calculate and interpret, and they recover the "true" scale values with higher accuracy in the presence of measurement error.

References

Budescu, D. V. (1984). Scaling binary comparison matrices: A comment on Narasimhan's proposal and other methods. Fuzzy Sets and Systems, 14, 187-192.

Gilliksen, H. (1959). Mathematical solutions for psychological problems. American Scientist, 47, 178-201.

Johnson, C. R., Beine, W. B., & Wang, T. Y. (1979). Right-left asymmetry in an eigenvector ranking procedure. Journal of Mathematical Psychology, 18, 61-64.

Lodge, M. (1981). Magnitude scaling — Quantitative measurement of opinion. Beverly Hills, CA: Sage Publications.

Marascuilo, L. A., & McSweeney, M. (1977). Nonparametric and distribution free methods for the social sciences. Monterey, CA: Brooks/Cole.

Nichols, W. G., & Gibbons, J. D. (1979). Parameter measures of skewness. Communications in statistics: Simulation and computation, B8, 161-167.

Ramsay, J. O. (1977). Maximum likelihood estimation in multidimensional scaling. Psychometrika, 42, 241-266.

Saaty, T. L. (1977). A scaling method for priorities in hierarchical structures. Journal of Mathematical Psychology, 15, 234-281.

Saaty, T. L. (1980). The analytic hierarchy process. New York: McGraw Hill.

Saaty, T. L. (1983). Foundations of the logical choice to harness complexity. The University of Pittsburgh: unpublished manuscript.

Saaty, T. L., & Bennett, J. P. (1977). A theory of analytical hierarchies applied to political candidacy. Behavioral Science, 33, 237-245.

Shepard, R. N. (1981). Psychological relations and psychophysical scales: On the status of "direct" psychophysical measurement. Journal of Mathematical Psychology, 24, 21-57.

Spetzler, C. S., & Stael von Holstein, C.A.S. (1975). Probability encoding in decision analysis. Management Science, 22, 340-358.

Stevens, J. C., Mack, J. D., & Stevens, S. S. (1960). Growth of sensation on seven continua as measured by force of handgrip. Journal of Experimental Psychology, 59, 60-67.

Stevens, S. S. (1946). On the theory of scales of measurement. Science, 103, 677-680.

Stevens, S. S. (1972). Psychophysics and social scaling. Morristown, NY: General Learning Press.

Stevens, S. S. (1975). Psychophysics: Introduction to its perceptual, neural, and social prospects. New York: Wiley.

Torgerson, W. S. (1958). Theory and methods of scaling. New York: Wiley.

Wallsten, T. S., & Budescu, D. V. (1983). Encoding subjective probabilities: A psychological and psychometric review. Management Science, 29, 151-173.

Williams, C., & Crawford, G. (1980). Analysis of subjective judgment matrices. Report R-2572-AF, Rand Corporation, Santa Monica, CA.

Wind, Y., & Saaty, T. L. (1980). Marketing applications of the analytic hierarchy process. Management Science, 26, 641-658.

Yager, R. R. (1979). An eigenvalue method of obtaining subjective probabilities. Behavioral Science, 24, 382-387.

Author Notes

This research was facilitated by Contract MDA 903-83-K-0347 from the U.S. Army Research Institute for the Behavioral and Social Sciences to the L. L. Thurstone Psychometric Laboratory, University of North Carolina at Chapel Hill. The views, opinions, and findings contained inthis paper are those of the authors and should not be construed as an official Department of the Army position, policy, or decision.

Dr. Budescu's work was partially supported by a grant from the Faculty of Social Sciences and Mathematics at the University of Haifa.