Guttman, L. A general nonmetric technique for finding the smallest coordinate space for a configuration of points. *Psychometrika*, 1968, *33*, 469–506.

Krantz, D. & Tversky, A. Similarity of rectangles: An analysis of subjective dimensions. *Journal of Mathematical Psychology*, 1975, *12*, 4–34.

Kruskal, J. B. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 1964, *29*, 1–27.

Kruskal, J. B. & Wish, M. *Multidimensional scaling*. Beverly Hills, Cal.: Sage Publications, 1978.

Lingoes, J. C. & Borg, I. CMDA-U: Confirmatory monotone distance analysis-unconditional. *Journal of Marketing Research*, 1978, *15*, 610–611.

Lingoes, J. C. & Roskam, E. E. A mathematical and empirical study of two multidimensional scaling algorithms. *Psychometrika*, 1973, *38*, Monograph Supplement.

Luenberger, D. G. *Introduction to linear and nonlinear programming*. Reading, Mass.: Addison-Wesley, 1973.

Roskam, E. E. A general system for nonmetric data anslysis. In J. C. Lingoes (Ed.) *Geometric representations of relational data*. Ann Arbor, Mich.: Mathesis Press, 1977.

Schönemann, P. H. Similarity of rectangles. *Journal of Mathematical Psychology*, 1977, *16*, 161–165.

Shepard, R. N. Representation of structure in similarity data: Problems and prospects. *Psychometrika*, 1974, *39*, 373–421.

Shepard, R. N. The circumplex and related topological manifolds in the study of perception. In S. Shye (Ed.) *Theory construction and data analysis in the behavioral sciences*. San Francisco: Jossey-Bass, 1978.

Skarabis, H. Multidimensional scaling with prescribed structures. In H. Skarabis & P. Sint (Eds.) *Compstat lectures 1*. Wien, Austria: Physika Verlag, 1978.

Manuscript received 8/28/78
First revision received 4/2/79
Final version received 8/9/79

FACTORIAL DEFINITIONS OF INTELLIGENCE: DUBIOUS LEGACY OF DOGMA IN DATA ANALYSIS

Peter H. Schönemann
Purdue University

1. EARLY DEFINITIONS OF "INTELLIGENCE": TOO MANY ANSWERS IN SEARCH OF A QUESTION

According to Cronbach (1949, p.101), "the outstanding success of scientific measurement of individual differences in behavior has been the intelligence test movement. Despite overenthusiasm and occasional errors that have attended their development, mental tests stand today as the most important single tool psychology has developed for the practical guidance of human affairs." Edwards echoes the same sentiment 25 years later: "If one were to choose the one area of psychology which is most representative of concerted effort and practical meaning, surely the measurement of intelligence and its implications would be foremost" (1974, p.3).

Vernon, on the other hand, has noticed some changes: "Since 1969 a considerable amount of research and critical writing has been published that does not consist merely of ideologically biased arguments" (1979). He is "disturbed by the current unpopularity of intelligence tests in the United States (and elsewhere) and ...therefore tried to analyze the reasons for this situation and how far it is justified." (1979, p.viii). One such critic is Blum (1978). After defining "pseudoscience... as a sustained process of false persuasion transacted by simulation or distortion of scientific inquiry and hypothesis testing", he concludes "the twentieth century's two clearest instances of grand pseudoscience have been Lysenkoism in the Soviet Union... and the various offshoots of Galton's theory in the United States and Great Britain, first under the title of eugenics and biometrics, and later referred to as psychometrics and behavior genetics" (p. 147).

As Wechsler discovered, it is much easier to define "intelligence tests" than it is to define "intelligence": "From the point of view of their avowed intent and wide use, intelligence tests are psychometric devices - in practice, sets of standardized questions and tasks, for assessing an individuals potential for purposive and useful behavior, at least in those aspects of it which one agrees to designate as intelligent. To be sure, there are many different definitions of intelligence, but nearly all intelligence scales appraise it in much the same way, namely by measuring a subject's mental abilities or current intellectual capacities." (Wechsler, 1974a, p. 30). Edwards (1974, p.13) tells us that Wechsler, at an earlier stage, was "content to describe general intelligence as a kind of energy which was neither definable nor measurable".

In his recent book, Bias in Mental Testing, Jensen (1980, p.170) taking a leaf from Spearman (1927, p. 11f), lists several verbal definitions of intelligence which have been offered over the years. Among these, Boring's probably comes closest to matching Wechsler's (above) in circularity: "Measurable intelligence is simply what the tests of intelligence test, until further scientific observation allows us to extend the definitions" (Boring, 1923). This version is sometimes offered as a paradigm of hardnosed operationalism. The trouble is that "intelligence tests" is used in the plural. Since the various intelligence tests do not correlate perfectly (see again Jensen, 1980, p. 314 for detailed numerical evidence), such a definition implies as many intelligences as there are intelligence tests. It was precisely this point which Spearman set out to deal with when he proposed his Two Factor Theory of intelligence. In this respect, and perhaps in others, he was ahead of his time, and perhaps not only his.

The historical connection between intelligence theories and factor theories is well known. Although generally most authors in the intelligence field make some reference to factor analysis, they differ in their commitment to factor theories as a basis for defining intelligence. At one time Wechsler (1939) seemed quite impressed with factor analysis: "More than 30 years ago, Professor Carl Spearman actually solved the problem by showing, through rigorous mathematical proof, that all intellectual abilities could be expressed as functions of two factors, one a general or intellectual factor common to every ability, and another a specific factor, specific to any particular ability and 'in every case different from that of all others'...We cannot enter into all this here, but can only indicate our own position by saying that Professor Spearman's generalized proof of the two-factor theory of human abilities constitutes one of the great discoveries of psychology" (1939, p.35). However, 20 years later, Wechsler had second thoughts: "...the profusion of factors discovered seems to contradict the intent or purpose of the factorial technique, the generally stated aim of which is to account for the major variance of a large battery of tests in terms of a minimal number of primary abilities or factors. Actually, there seem to be more factors than available tests, certainly good tests of intelligence" (1958, p.127).

Vernon, another prolific writer on intelligence, is equally ambivalent about the value of factor analysis for intelligence theory. On the one hand, he defends factor analysis and takes issue with "critics [who] deny the relevance of factor analysis to defining intelligence because of the disagreement between different factor models. But I have tried to show that these discrepancies are exaggerated" (1979, p.61; my emphasis). In a similar vein, on p.67: "The apparent conflict between the various models does not imply that factor analysis is worthless. Discrepancies arise mainly because heterogeneous samples are used in some studies, and selected, homogeneous groups in others." On the other hand, "we can see that factor analysis does not yield any definite solution to the problem of uni- or multi-dimensionality of intelligence, though there is more agreement than appears on the surface when account is taken of the effects of the heterogeneity of the population" (p. 61). In 1970 (p. 105), he added cryptically: "Though I cannot offer any satisfactory answer, it is some consolation to recollect that they [the various factor solutions] are mathematically inter-convertible - in other words, that factor analysis by itself cannot tell us how man's traits and abilities are organized."

Jensen, by comparison, seems to have less tolerance for ambiguity. He is firmly committed to factor analysis as a basis for intelligence theory: "The basic vocabulary, concepts, and typical results of factor analysis are so absolutely central to any theoretical understanding of intelligence and to the scientific definition of the concept itself that the reader who does not possess some idea of how factor analysis works will simply run the risk of reading the whole remainder of this book with misapprehension every time certain key terms are used, including the key concept of 'intelligence'" (1980, p. 185, my emphasis). Jensen has become the center of some controversy because of his unorthodox views on the inheritance of intelligence (Jensen, 1969). Vernon seems to think this controversy could have been avoided had Jensen demonstrated a better sense of timing: "It was indeed unfortunate that publication of the 1969 article coincided with the peak of college student activism against the establishment' and the pressures for black power" (1979, p. 14). Rather than arguing about Jensen's conclusions, I propose we now examine his basic premise: can factor analysis furnish an operational definition of intelligence?

2. SPEARMAN: A NEW QUESTION, WITH A NEW ANSWER

"Chaos itself can go no further! The disagreement between different testers - indeed even between the doctrine and practice of the selfsame tester - has reached its apogee. If they still tolerate each other's proceedings, this is only rendered possible by the ostrich-like policy of not looking facts in the face. In truth, 'intelligence' has become a mere vocal sound, a word with so many meanings that finally it has none...test results and numerical tables are further accumulated; consequent action affecting the welfare of thousands of persons is proposed, and even taken, on the grounds of - nobody knows what!" (Spearman, 1927, p.15).

On reading the first few chapters of Spearman's Abilities of Man (1927), one quickly discovers that Spearman was thoroughly disillusioned with the semantic confusion surrounding him. By placing the definition of intelligence on a more solid empirical foundation, he hoped to make it more fruitful for empirical research.

He began by taking stock of three major "doctrines" of intelligence which were current at the time:

(i) the monarchic doctrine "...which assumes mental ability to lie under the sovereign rule of one great power named 'intelligence'" (p. 4).

(ii) the oligarchic doctrine "...of several different powers (abilities, capacities, levels, or however else they may be named). Typical instances are 'judgment', 'memory', 'invention', 'attention'. Each of these is taken to constitute a separate function or behavior-unit on its own account. Accordingly, each allows and requires its own separate measurement" (p. 26).

(iii) the anarchic doctrine: The "crude view that all abilities are independent" (p.55) which he traces back to Thorndike (1903, p.39): "the mind is a host of highly particularized and independent faculties". Spearman rejects this "crude view" even if it were amended to allow the many different abilities to be more or less correlated.

Upon closer scrutiny, Spearman rejects all three "doctrines" and supplies his own, fourth:

(iv) the eclectic doctrine of two factors, one general and one specific factor for each test, as Wechsler has outlined it in the above (1939) quote.

In order to appreciate the difference between Spearman's efforts and those of his predecessors, we must now take a closer look at his "Two Factor Theory" (TFT) of intelligence. There are basically two major differences. (i) Spearman developed his theory from a specific, replicable empirical observation of some generality; the observation that all intelligence tests he knew were positively intercorrelated ("law of the positive manifold") and, (ii) he stated his theory in mathematical terms, so that its logical structure could be more easily checked for stringency than the vague verbalisms of his predecessors. Both steps took a measure of courage and commitment to traditional scientific values: the empirical observation he started with might at a later time be shown to be false, and the logical structure he proposed to describe it with might later be shown to be inconsistent. He, therefore, made it relatively easy for his rivals to attack him. These risks are considerably lessened if one employs a definition which contains sufficiently many undefined terms to be virtually meaningless, or which is logically contradictory, e.g. by defining "intelligence" as something which cannot be defined. Logical contradictions are especially useful for constructing irrefutable theories, because they imply any statement whatever.

Spearman proposed a definition of "intelligence" which was part of an empirically falsifiable theory. This theory was intended to "explain" a "law", namely the "law of the positive manifold", which Guttman now (1977, p.7) calls "the first law of intelligence": "This [i.e. the eclectic] doctrine was based upon what we have all along been finding of such paramount importance, namely, the correlations between abilities" (Spearman, 1927, p.72).

Spearman explained the positive manifold by assuming that each test y_i is composed of exactly two components

(2.1) $y_i = \hat{y}_i + e_i$

which are uncorrelated with each other. The first component, $\hat{y}_i$, is proportional to a "general factor" x (Spearman called it "g", for "general ability") which is "common" to all the tests y_i. The second component, e_i, is proportional to a second factor z_i which is not only uncorrelated with x, but also with every other z_j, and thus, "specific" to the i'th test (hence: TFT). If one defines the two constants of proportionality a_i and u_i, respectively, then the TFT can be written

(2.2) $y_i = a_i x + u_i z_i$

or, in matrix notation,

(2.3) $\eta = \alpha x + U \zeta$

where $\eta'=(y_1,\dots,y_p)$ contains the p observed intelligence tests, x is the common factor ("g"), and $\zeta'=(z_1,\dots,z_p)$ contains the p specific factors. The vector $\alpha'=(a_1,\dots,a_p)$ contains the weights with which the general ability x enters each intelligence test y_i, and the diagonal matrix $U=diag(u_1,\dots,u_p)$ contains the weights with which each specific factor enters the i'th test. Both sets of weights were assumed to be positive, i.e. the general factor and the specific factors entered each test to various degrees, but always with the same sign. This explained the positive manifold. To complete the formal statement of Spearman's model, we have to add to the "structural equation" (2.3) the covariance stipulations which Spearman imposed:

$$(2.4)\ \mathrm{var}(x,\ \zeta') = I_{p+1}.$$

This equation simply states that all "latent variables", the common factor x, and the p specific factors z_p, are uncorrelated with each other (the variances being set to one arbitrarily). This covariance condition is an essential part of Spearman's model. It is of fundamental importance if one one wants to understand its formal properties and limitations. Most texts have been somewhat delinquent in stating this model explicitly and completely. This has led to considerable confusion which could have been avoided.

It is clear that Spearman's TFT goes beyond a simple restatement of the "law of positive manifold" it is supposed to help explain, because it has other, additional consequences, which can be empirically tested. If one denotes the covariance matrix of the observed variables y_i by $\mathrm{var}(\eta)=\Sigma$, then Spearman's model implies

$$(2.5)\ \Sigma = \alpha\ \alpha' + U^2.$$

Since all the regression weights a_i in α are assumed to be positive, a reordering of the observed variables y_i according to the magnitude of the a_i will produce a correlation matrix with a "hierarchical" structure: all correlations will taper off from the upper left hand corner of the permuted table. However, since the same is true for the manifest correlations of entirely different structures, such a test by inspection could be misleading. A stronger test is provided by the "ideal rank" of Σ. Since $\alpha\alpha'$ is of rank one, and U^2 affects only the diagonal elements of Σ, one finds that all 2x2 determinants of Σ which do not involve the diagonal elements (and which Spearman called "tetrad differences"), must vanish. He spent considerable effort on developing statistical tests for this prediction of the model. An equally obvious implication of (2.5) is that the columns of Σ, $a_j\alpha$, are proportional to each other, if one again omits the rows corresponding to diagonal elements. This gives a second, somewhat cruder test of the model, which is more easily applied in practice: "the average intercolumnar correlation" of Σ should be near unity, if the model holds. Thus, Spearman's approach was thoroughly "confirmatory", to use another fashionable term. This should not come as a surprise, because he had intended the factor model as a mathematical theory for intelligence, not as the device for "exploratory data reduction" which it later became.

On applying these tests to available intelligence data, Spearman found his theory confirmed: "The average intercolumnar correlation from tables of 14 different investigators, summarizing 30 years of psychological researches and representing a great wealth of test material, was unity, as predicted by the unifocal hypothesis of a general factor. It seemed to be the most striking quantitative fact in the history of psychology" (Dodd, 1928, p.214). It is therefore understandable that Spearman should have felt elated: "Indeed, so many possibilities suggest themselves that it is difficult to speak freely without seeming extravagant...It seems even possible to anticipate the day when there will be yearly official registration of the 'intellective index', as we will call it, of every child throughout the kingdom...The present difficulties of picking out the abler children for more advanced education, and the 'mentally defective' children for less advanced, would vanish in the solution of the more general problem of adapting education for all...Citizens, instead of choosing their career at almost blind hazard, will undertake just the professions really suited to their capacities. One can even conceive the establishment of a minimum index to qualify for parliamentary vote, and above all for the right to have offspring" (Hart and Spearman, 1912, pp. 78-79).

3. THOMSON: ANOTHER ANSWER TO THE SAME QUESTION

After a few years of consolidation, a number of questions were raised about this new intelligence theory. More interesting than the questions is the way in which they were eventually answered. Three of these questions are of particular importance for any attempt to define "intelligence" operationally through factor analysis:

(i) It turned out that the model did not always fit as well as Spearman had hoped. In many cases the tetrad differences were not zero, and the average intercolumnar correlation was not unity. Thus, there might be more than one intelligence factor.

(ii) In (1916), Thomson showed that he could produce correlation matrices with all the properties predicted by Spearman's theory, although the data had been generated from a very large number of common factors, instead of only one general factor. Thus, there might be infinitely many intelligence factors.

(iii) In 1928, Wilson showed, in his review of Spearman's "Abilities of Man", that the Two-Factor model had a built in flaw which threatened the very purpose it was designed to achieve: g, i.e., the common factor x in eq. (2.3), and the specific factors z_j, were not uniquely defined by the observed variables y_i even when the model fit the data perfectly. Thus, there might be no operationally definable intelligence factor at all.

Technically, the first problem is easily dealt with by simply generalizing Spearman's model from one to m ($1 \leq m < p-1$) common factors. Spearman was reluctant to adopt this easy solution, presumably because this would have meant abandoning his "eclectic doctrine" in favor of the "oligarchic doctrine" which he found unacceptable on philosophical grounds (1927, p.39).

The technical aspects of this extension had been discussed by Garnett (1919) and Dodd (1928), some time before Thurstone established his claim on this "multiple factor analysis" generalization (1931, 1947, p.vi) of Spearman's model. The matrix formulation of this more general model is

(3.1) $\eta = A\ \xi + U\ \zeta$

(3.2) $var(\xi') = \Psi,\ var(\zeta) = I_p,\ cov(\xi, \zeta) = \phi,$

where A is a pxm (common) "factor pattern", $\xi' = (x_1, \ldots, x_m)$ contains the m common factors, and Ψ is their intercorrelation matrix (which is often taken as I_m). To avoid confusion with other factor models, I shall call (3.1), (3.2) the "Conventional Factor Model" (CFM), because it is the factor model which has been assumed, implicitly or explicity, in virtually all texts on factor analysis since the thirties (see Schönemann, 1978, for a long list of such texts). The definitions (3.1), (3.2) imply at the covariance level

(3.3) $\Sigma = A\ \Psi\ A' + U^2,$

which corresponds to (2.5).

This resolution of problem (i) creates two new problems, the "number of factors problem" and the "rotation problem", which will be taken up in later sections. The main purpose for introducing the multiple factor generalization at this point is to have the slightly more general notation (3.1), (3.2), available when needed for the discussion of the other two problems, which profoundly affect the theoretical and psychological interpretation of the factor model.

In (1916), at the height of WWI, G. Thomson, "after correspondence with Major Spearman" (p.271), published a

paper he had already drawn up in 1914. It would be understandable if Major Spearman had been less in a hurry than usual to see this particular paper in print, because Thomson demonstrated, with what we would now call a "Monte Carlo experiment", "...that a certain set of correlation coefficients, which we <u>know</u> to contain no General Factor, would be claimed by Professor Spearman as giving further support to the existence of such a factor. <u>There is therefore nothing to show whether the many cases brought forward by him really contain a general factor or not</u> (p. 280, emphasis in the original).

To lend substance to this heresy, Thomson "threw in all 5220 dice, in 36 groups of 145 each, to represent ten tests in a class of 36 boys" (p. 278), in accordance with the following model:

He postulated a large number, say m>>p, "group factors" x_k which are uncorrelated with each other and have unit variance. Each observed test

(3.4) $y_i = \Sigma\ x_k,$

(where k ranges over a subset of the index set), is the sum of a random sample of size $m_i = mp_i$ of the m x_k. The samples defining the i'th and j'th test are drawn independently. This theory is evidently a close relative of Thorndike's "anarchic" Theory of Bonds. Let us write this model

(3.5) $\eta = S\ \xi,$

where $\xi' = (x_1, \ldots, x_m)$ contains all m factors, and m is large relative to p, the number of observed variables y_i. The rows σ_i' of the pxm matrix S contain $m_i = mp_i$ ones and $m - m_i$ zeroes in random distribution, i.e. they have the form

(3.6) $\sigma_i' = \sqrt{p_i}\ (0,0,1,0,1,1,0,\ldots,1,0),$

(see Appendix A1.1 for an illustration). The covariance assumption is

(3.7) $var(\xi) = I_m,\ m >> p$

so that

(3.8) $var(\eta) = \Sigma = S\ S',$

in apparent contrast to (2.5) (see Appendix A1.2). However, the correlation between the i'th and the j'th test turns out to be, in view of (3.7)

(3.9) $corr(y_i,\ y_j) = \sigma_i'\sigma_j / \sqrt{\sigma_i'\sigma_i\ \sigma_j'\sigma_j}.$

As m increases, the numerator will approach mp_ip_j as a consequence of the independence assumption. Hence, in the limit, the correlation will be

$$(3.10)\ \rho_{ij} = mp_ip_j/\sqrt{(mp_i)(mp_j)} = \sqrt{p_ip_j}.$$

If one defines $\pi'=(\sqrt{p_1},\ldots,\sqrt{p_p})$, var($\eta$) becomes

$$(3.11)\ \Sigma = \pi\,\pi' + (I\text{-diag}(\pi\pi'))$$

which is identical to the covariance matrix (2.5) predicted by Spearman's model.

Moreover, it is well known that MM'=NN' implies that M and N relate by a partial isometry, which can be completed into an orthogonal matrix T by adjoining as many zero columns to one of the two matrices until M and N have the same number of columns. Hence

$$(3.12)\ S\ T = (\alpha,\ U,\ \phi),\ T'T = I_m,$$

for some mxm orthogonal T. This means either solution can be carried into the other, as Garnett observed in (1920).

Thus, although the psychological assumptions behind both theories are in complete disagreement with each other, the predictions they make about Σ are exactly the same (at least in the limit, as m goes to infinity, a condition which Thomson did not insist on. For "large" but finite m, var(η)=Σ has a tendency towards hierarchy and low ideal rank). Finally, the two sets of "intelligence factors" can be converted into each other.

Does this mathematical equivalence mean that both psychological theories are interchangeable? Spearman thought it did. He naturally preferred his own theory and argued it should be retained because it was more "parsimoneous". On the other hand, Thomson (1920), and later Mackie (1928), strongly objected to this interpretation with arguments which, so far as I know, have never been convincingly refuted.

This is how Thomson (1920, p.319) interpreted his results: "Hierarchical order [i.e. ideal rank one] will arise among correlation coefficients unless we take pains to suppress it. It does not point to the presence of a general factor, nor can it be made the touchstone for any particular form of hypothesis, for it occurs even if we make only the negative assumption that <u>we do not know</u> how the correlations are caused, if we assume only that the connexions are random" (his emphasis). To choose between both theories on the basis of the parsimony principle (which under Thurstone assumed the status of an unquestioned dogma, as we shall see) was unacceptable to Thomson: "If I make the discovery

that the angles of a quadrilateral are equal in sum to four right angles, I may not conclude that it is a square. This is, by analogy, what Prof. Spearman did when he noticed hierarchical order, and deduced a General Factor. True, the angles of a square are equal to four right angles, and the thing may be a square: and similarly a General Factor may exist. But the marks of a genus do not define the species. <u>Nor may I conclude that I may call it a square because that is simpler</u>" (p.321, my emphasis).

This distinction between mathematical equivalence of two models, on the one hand, and lack of equivalence of the substantive theories they are supposed to model, on the other, was further sharpened by Mackie (1928). His argument pivots on the observation that Spearman's "g" must correspond to all of ξ in Thomson's model, since the correlation of y_i with g is given by $a_i = \sqrt{p_i}$, and p_i is the proportion m_i/m of factors x_k sampled by the i'th test y_i. This creates the dilemma of having to explain where the specific factors come from, after Thomson's factors have been mathematically transformed into Spearman's "g".

Although the quote from Mackie (1928) is a bit longer than usual, there are at least three reasons for reproducing it in full. Firstly, I consider the point at issue as of some importance not only for factor theory and intelligence theory, but also for psychological modelling in general. Secondly, it would be impossible for me to improve on Mackie's statement of his case:

"We may venture to make our argument clearer by an analogy. Suppose that some archaeologists find the ruins of three temples, one of which had been made of brick and stone, another of brick and wood, and the third of stone and wood. Some of the archaeologists might say that in the building of each temple we see the operation of two factors, one of which is common to all: this might be the religious urge of the nation which built them. The factors peculiar to each of the temples might be respectively: (1) Absence of trees in the locality, (2) want of stone, and (3) abundance of stone and wood. Others of the archaeologists say that the nation in question possessed skilled builders and that the three temples are samples of their work, the first being due to brickmakers and masons, the second to brickmakers and carpenters, and the third to masons and carpeters. The first school of archaeologists then claims that the second theory is but a variant of theirs, for, they say, we can make a mathematical transformation by means of which the brickmakers, masons and carpenters are replaced by four influences, one affecting all three temples and the other three one each. If it is inquired what those influences are, the reply must be: The general influence is the work of the whole set of builders; the specific influence in the first temple is the work of the brickmakers and masons hindered by the carpenters, and similarly for the others. A sort of mathematical equation would be:

Brick-and-stone Temple = (Work of brickmakers, masons and carpenters working with part of their might) + (Work of brickmakers and masons working with the rest of their might − Work of carpenters already done).

It would then be open to the second set of archaeologists to point out that this was just what they said at first, except that the carpenters had been brought in to build and then knock down what they had built; and that they might justifiably prefer to think that the carpenters were never there at all. As long as they chose to keep their own hypothesis about the group of builders it would be true to say that their theory could not be regarded as variant of the other one.

We conclude then that it is only in the most formal mathematical sense that the Sampling Theory can be brought in under the Two Factor Theory; that if we adhere to the hypothesis underlying the sampling theory the interpretations we are compelled to put upon the specific factors obtained by the mathematical transformation is such as to show that these factors are mere mathematical fictions. Per contra, if, having arrived at the Two Factor Theory, we make the transformation from it, then the elements of the Sampling Theory expressed by means of the transformation are mere mathematical fictions. If either theory should be abandoned, it is not because they are equivalent" (p. 620f).

The third reason for devoting so much space to Mackie is a recent reference to him by Maxwell (1972), who seems to have misunderstood the very point Mackie tried to make. Maxwell, a well-known statistician, has lately become disenchanted with "the proliferation of factors" (p.9) in the Spearman/Thurstone sense, after having devoted some of his efforts to developing methods for "estimating" them (e.g., Lawley and Maxwell, 1972, ch. 8). He now favors Thomson's theory, and therefore is in need of a new justification for "estimating factor scores" (p. 7). After converting Thomson's solution "by simple linear regression" into Spearman's, he warns us not to interpret the residual as a specific factor: "According to Thomson's sampling theory, such an interpretation, as Mackie (1928) has shown, would be inconsistent, *and it is best to think of the variate e_i merely as a contrast between the mean value of the component sampled by [y_i] in a population of people and the average, g, of all their components*" (p.5, my emphasis).

Now it seems to me such "regression estimates" and "contrasts" are exactly what Mackie must have had in mind when he spoke of "mere mathematical fictions". The last

thing he would want us to do is to estimate them. So far as I can see, there are no factors left to be estimated once one adopts the Thomson interpretation: before the transformation there is an unknown large number of factors, and after the transformation into the corresponding Spearman factors, we are left with a smaller number of mathematical fictions. What we can, and perhaps sometimes should do, is compute scores on linear combinations of the tests. However, such "components" (Schönemann and Steiger, 1976), are neither factors nor "factor estimates" by any reasonable interpretation of the term. Let us now ask whether it makes sense "to estimate factor scores" in the CFM of Spearman and Thurstone:

4. WILSON: QUESTIONS SOME OF THE ANSWERS

When E.B. Wilson, in 1928, reviewed Spearman's "Abilities of Man" for Science, he already had a suffiently long and distinguished career behind him (Hunsaker and Mac Lane, 1973), that he could afford the luxury of admitting interest in psychology without arousing suspicions among his peers. Spearman greeted Wilson with befitting deference: "If any event is more likely than another to quicken the progress of psychological mathematics, it is the entry on the scene of a mathematician so eminent and so free from prejudice as Professor E.B Wilson" (1929, p.212).

Wilson did not dwell on occasional lapses of rigor which so often fascinate his modern successors. Instead, he preferred to look at the larger picture: "Science advances not so much by the completeness or elegance of its mathematics as by the significance of its facts. You cannot upset the findings of the "Origin of Species" either by the contraposition of your religious convictions or by observing that Darwin's statistical technique was not up to standard" (p.244). Having previously (1903) tangled with Hilbert about "The so-called foundations of geometry", he was careful, on this occasion, to be quite concrete: "Try a case. Let the marks of 6 students on 3 tests be (the first columns give the actual marks, the second columns give the differences from the means)..." (p.246).

However, it is not certain that Spearman welcomed this degree of explicitness because, "what we have shown is that the complete solution [for exact factor scores implied by the given "marks"] can be obtained but is indeterminate. We have no need of harder mathematics than the solution of a set of linear equations and 1 quadratic equation...It would be quite superfluous to introduce this higher mathematics, involving a probability theory which probably does not apply anyhow, to make determinate (if it does) that which without it seems indeterminate" (Wilson, 1928, p.246).

In retrospect it may seem odd that it should have taken almost 25 years before someone noticed that the model Spearman had proposed in 1904, and which had been extensively debated by numerous competent people since then, has a built in indeterminacy which threatens to undermine the very purpose the model was to serve, "...the yearly official registration of the 'intellective index'" which Spearman had so eagerly anticipated and which has become an established custom by now. If this index is not uniquely definable after all, how then can a responsible person "even conceive the establishment of a minimum index to qualify for parliamentary vote, and above all for the right to have offspring"? And what has factor analysis then contributed towards a more unambiguous, "scientific", definition of "intelligence", which had been Spearman's main aim?

In order to treat this question in sufficient generality to encompass the modern factor theories of intelligence, I shall state it in the context of multiple factor theory. A clue that not all is well with the CFM is given by the partial covariance matrix of the common factors, ξ, after all the observed information in η has been partialled out again. This partial covariance matrix,

$$(4.1)\quad \mathrm{var}(\xi|\eta) = \mathrm{var}(\xi) - \mathrm{cov}(\xi,\eta)\mathrm{var}^{-1}(\eta)\mathrm{cov}(\eta,\xi)$$

$$= \Psi - \Psi A'\Sigma^{-1}A\Psi \neq \phi,$$

is not zero for the conventional factor model (3.1), (3.2). This raises the question where the implied residual variables come from which unexpectedly emerge in (4.1). The answer is that these additional variables have to be pulled literally out of a hat: they must be randomly chosen in such a way that they are uncorrelated with the observed test variables y_i in η, so that the latent variables ξ, ζ can satisfy (3.1), (3.2) for a given η. In the most general technical terms, the answer has been stated by Guttman (1955) for the correlated multiple factor case: in order that the vector of common factors, ξ, and the vector of specific (unique) factors, ζ', satisfy the CFM (3.1), (3.2), for a given observed test vector η, and (arbitrarily chosen) nonsingular $\mathrm{var}(\xi)=\Psi$, they must be constructed as

$$(4.2)\quad \xi = \xi_1 + \xi_2;\ \zeta = \zeta_1 + \zeta_2,$$

where ξ_1 and ζ_1 are linear functions of the observed tests y_i (and, thus, functions of the observed testscores y_{ij} in the sample. See Guttman, 1955, or Schönemann and Wang, 1972, eqs. (2.12), (2.13) for the defining weights). The vectors ξ_2, and ζ_2, on the other hand, are linear functions of a second set of m independent random variables $\sigma'=(s_1,\ldots,s_m)$ which are uncorrelated with η, and which satisfy the covariance conditions

$$(4.3)\quad \mathrm{var}(\sigma) = I_m,\ \mathrm{cov}(\eta,\sigma) = \phi.$$

These constraints leave a great deal of freedom for choosing σ and, consequently, for defining the "factors" ξ, and ζ, given η, and hence these factors are "indeterminate". An obvious question to ask is how indeterminate they are in practice. To answer it, one needs a measure of the indeterminacy. Guttman (1955) proposed the "minimum correlation between equivalent factors",

$$(4.4)\quad \rho_k = \mathrm{corr}(x_k,\ x_k^*) = 2\gamma_k'\Sigma^{-1}\gamma_k - 1,$$

where γ_k is the k'th column of the factor structure $G=A\Psi$, for assessing the extent of the indeterminacy. This index is simply the correlation between the k'th common factors of a pair of two equivalent solutions ξ, ξ^* which have been selected so as to be minimally correlated among all possible solutions. This index varies in theory between −1 and 1. Schönemann and Wang (1972) found that it varies over this range also in practice. This means that in many cases the factors are very poorly defined by the observed vector η. In the intelligence domain, the minimum correlation will usually not be negative, and it may be reasonably high as long as not too many factors are "retained". However, as it becomes necessary to extract more factors, e.g., to satisfy a statistical criterion of fit, then the last few factors may be very poorly defined. A realistic illustration of what can be expected for a currently popular intelligence test, the WISC, is given in Appendix A3.4.

In order to illustrate the indeterminacy issue as concretely as possible, a complete factor analysis of the Wechsler Adult Intelligence Scale (WAIS) is presented in Appendix A2, together with the results of a regression component analysis of the same data. Following convention, the original correlations were factored for two common factors, which were rotated (obliquely) to simple structure. The resulting Verbal (V) and Performance (P) factors are highly correlated (.76). Although the chi-square test rejects the CFM for only two factors and would require five factors for a statistical fit for these data, the more conventional two factor solution is preferred here for the sake of illustration. The introduction of more factors would create more new problems than it would solve, including a loss of estimability of the factor pattern itself. The dubious merits of the chi-square test will be discussed in section 6. The estimated factor patterns, A, U, and the factor intercorrelations Ψ were used to construct 20 "observed scores" y_{ij} which reproduce the estimated covariance matrix $\Sigma=A\Psi A'+U^2$ exactly (A3.2). Two sets of minimally correlated equivalent factor scores, (X, Z), and (X*, Z*), which, together with (A, U) reproduce Y equally well, and which both have the same intercorrelation matrix (Ψ for the common factors, and I for the specifics) are given in A3.1. To simplify the interpretation of the (intelligence) factor score discrepancies in X and X*, both were converted into IQ's (A3.3, setting μ=100, σ=15). Although the minimum correlations for both factors are fairly high (.89, .76), there are noticeable discrepancies, especially between the two sets of equivalent "performance IQ's" (2nd and 4th column in A3.3). These (underlined) discrepancies are not due to errors of estimates in the usual sense, because they would remain even if Y were to remain perfectly stable upon replication. The two columns labelled x_1, x_2, or, equivalently, the next two columns labelled x_1*, x_2* are the exact intelligence factor scores for the 20 ficticious subjects, by the definition of the CFM. Hence it is quite misleading to say "factorscores have to be estimated because they cannot be computed exactly". Both sets X, X*, are exact.

It stands to reason that only after the algebraic facts of the indeterminacy problem are clearly understood does it make sense to try to understand what they mean. Regrettably, these simple facts have been carefully hidden from the view of the unsuspecting users of factor analysis by virtually all textbook writers since the thirties, beginning with Thurstone (1938). Two rare exceptions are Thomson (1951) and, eventually, Mulaik (1972). The very influential texts by Harman (1960, 1967), which dominated the field in the post-Thurstone period, together with the more applied texts by Fruchter (1954, 1964), Cattell (1952) and others, never mention the indeterminacy problem at all. This is remarkable in view of the fact that it had become the focus of a lively debate all through the thirties (see Steiger and Schönemann, 1978, for a historical review), which included a number of well-known and established scholars of unquestioned technical competence, among them Wilson, Spearman, Thomson, Heywood, Piaggio, Irwin, Camp, Ledermann, and, eventually, Kestelman and Guttman - but never Thurstone, so far as I know.

That the subsequent collective amnesia was not entirely accidental has been demonstrated repeatedly, for example when Steiger and I tried to publish one of the implications of the indeterminacy which had not been noticed before:

The factors x_k, z_i of the CFM can always be chosen so as to perfectly predict any criterion whatever, regardless how it correlates with the observed variables y_i.

The proof of this curiosity is quite simple (for details and extensions see Schönemann and Steiger, 1978; Steiger, 1979): let the factor analysis be based on a set of intelligence testscores y_{ij} for N pupils, and assume, for convenience, that Spearman's TFT fits these data perfectly for one general factor g. Now suppose we wish to predict the shoesize of the fathers from the pupil's intelligence

scores g_i (x_i in the notation of (3.1), (3.2)) and the unique factor scores z_{ij} jointly. Let this criterion be w. After the part of w, if any, which correlates with the determinate part (ξ_i in (4.2)) of g has been removed, the residual will be uncorrelated with the observed scores y_{ij}. Hence we can use it as the additional random variable s needed to define g, ζ, so that the multiple correlation of shoesize with the intelligence factor g and the specific factors z_i will be unity. When we submitted this result in a short Note to the British Journal of Mathematical and Statistical Psychology in 1976, the editor, despite his best efforts over a period of 23 months, was unable to reach a decision because he had "failed to find anyone willing to give you 'technical reviews' (your letter March 13th). The brief comments I have question 'motivation' or 'what is perceived to be the problem' rather than your algebra" (BJMSP, 1978).

Spearman, in his time, tried to meet Wilson's challenge with some rational argument. He was quick to notice that the problem would "vanish in the limit" if it were possible to lengthen the testbattery indefinitely without incurring further common factors. As Guttman (1955) has shown, this condition can be further weakened, in the multiple factor case, by requiring only that the ratio m/p of the number of common factor relative to the number of observed variables approach zero in the limit as p goes to infinity. A very simple argument which shows this was given by Schönemann (1971), who found that the "average minimum correlation" τ obtained upon averaging the correlations (4.4) for the common factors together with the analogous minimum correlations for the unique factors z_i

(4.5) $\tau = (p-m)/(p+m) = (1-m/p)/(1+m/p)$,

is a constant which does not depend on the data, Y, A, or U, at all. It, therefore, can be computed as soon as m, the number of common factors is known. It is immediately obvious from (4.5) that this average, and hence all minimum correlations, will go to one as m/p goes to zero.

However, this is more easily said than done. Not only because it presupposes tireless patience on the part of the subjects who take the tests, but also because one often finds that adding more tests adds more factors. Moreover, as Mulaik and McDonald have recently (1978) shown, even if it were possible to find a suffiently large number of tests to suppress the indeterminacy without incurring further factors, the eventual limit will depend on the tests which are added. Investigator A, and investigator B, starting with precisely the same battery of p tests y_j, will end up with two quite different determinate "g"'s. Hence further conditions have to be added before the indeterminacy problem can be made to disappear "in the limit" on paper.

Following up on earlier groundbreaking work by McDonald (1972, 1974), Williams (1978) reinvestigated these "stability conditions" in a paper entitled "A definition for the common- factor analysis model and the elimination of the problems of factor score indeterminacy" (see Kruskal, 1976; Guttman, 1975; Schönemann, 1978, for some muffled echoes on McDonald's earlier attempt to eliminate this problem). Williams proposes a revised formulation of the CFM which includes these conditions as part of the definition of "the" factor model as a stochastic process which, he hopes, may eventually converge on a set of uniquely defined factors as the battery is lengthened. After dismissing earlier formulations of the factor model, e.g. the CFM, as "trivial" (p. 295; see also McDonald, 1972), Williams arrives at a number of intriguing conclusions which should provide food for thought for the intelligence community: "If the space of all random variates to be considered in the definition of factor scores is essentially contained in the span of X_i, $i=1,2,\ldots$, then either a model exists with essentially unique scores or it does not exist" (p. 300). He explains our failure to have noticed any of this before with the theory that, (i) "no adequate model has ever been set out before" and (ii) "the concept of a random variable was not understood well enough" (p. 303).

This explanation does not change the fact that the CFM has been used for several decades by legions of research workers, nor does is tell us what to do with the dubious legacy of empirical and theoretical results they have left behind, including, as it were, most extant intelligence theories. The practical implications of these results may not always have been "trivial" for those whose careers were decided on the basis of theoretical premises which now are lightly dismissed as no longer valid. Another weakness of Williams' theory is that it does not adequately explain why many noted statisticians, e.g., Lawley, Maxwell, Jöreskog, Anderson, Rao, to name a few, had overlooked the indeterminacy for decades, just like their psychological peers. Many of them worried how best to estimate (the indeterminate) factor scores, and the regression weights for factors which have the absurd properties Schönemann and Steiger were not supposed to make public. And lastly, such an explanation is somewhat unfair to E. B. Wilson, whom Williams does not mention. Wilson knew the meaning of a "random variable" as well as anyone: he had anticipated the modern formulation of a "confidence interval" (1927), and he had served as President of the American Statistical Association in 1929.

Williams does not reveal how he intends to test his stability constraints with fallible data. It is therefore difficult to know how he found out that "factor analysis studies usually involve many more variates than common factors of interest". It would be most fortunate if his conjecture were true, because "these results would appear to indicate that stable sets of factor loadings can easily be identified in practice" (p. 306). Williams' approach has the further advantage of completely circumventing the nagging power question, to which I shall return in section 6: what is the probability that we accept the factor model when in fact it is false? If it should turn out that this problem is more serious than its widespread neglect suggests, what then? Are we then to continue Spearman's recommendation to decide on the basis of his intellective index who may have off-spring or not?

Like Williams, Mulaik and McDonald (1978) appear to be unperturbed about the potentially adverse social consequences erroneous intelligence theories may have (see, e.g., Kamin, 1974, ch. 1): "Thus with a well-defined behavior domain, researchers should be able to treat the hypothesis that the domain has a determinate factor space as a null hypothesis to be accepted until data inconsistent with it is obtained" (p. 192).

5. THURSTONE: SIMPLE ANSWERS FOR MOST QUESTIONS

"Twenty years have now elapsed since Professor Thurstone's ingenuity pulled the factor problem out of its tetrad difference quagmire. Most of us have watched Thurstone's brainchild grow...Indeed, even as an infant, it was forced to throw aside its swaddling clothes in order to kick back at the Spearmans, the Tryons, the Anastasis and others who sometimes rightly, oft-times wrongly, pointed a finger at supposed imperfections. The youngster weathered more or less successfully the many storms it had stirred up...and was ready for the sober coming of age evaluation of its strengths and weaknesses by Dael Wolfle" (McNemar, 1951, p. 353).

As this quotation suggests, the intervention of Thurstone in the early thirties marks a major turning point in the history of factor theory:

(i) He popularized the multiple factor generalization of Spearman's TFT for the description of intelligence and advocated its use also in other content areas "as a general scientific method" (1947, p.55), rather than viewing it as a mathematical theory for intelligence, as Spearman had originally intended it.

(ii) He made numerous algorithmic contributions to simplify the then arduous task of performing a factor analysis. He thus paved the way towards the analysis of much larger data sets than had been previously possible.

(iii) He contributed the criterion of "Simple Structure" as a guiding principle for resolving the rotational indeterminacy of the CFM when $m>1$: since, in this case,

$$(5.1)\quad \Sigma=AA'+U^2=A^*A^{*\prime}+U^2, \text{ with } A^*=AT \text{ for all orthogonal } T,$$

some such principle is needed to select the factor pattern A when there are more than one common factor: "The simple structure concept was invented, if anything so simple can be called an 'invention', as a compromise with this problem. It was the simple idea of finding the smallest number of parameters for describing each test. In numerical terms this is a demand for the smallest number of non-vanishing entries in each row of the oblique factor matrix...This simple idea has turned out to be a much more powerful analytical device than was at first anticipated" (p. 333, his emphasis).

(iv) He developed a theory of "higher order factors" as a natural extension of this simple structure principle: in many cases (which include intelligence test data) simple structure can be improved by admitting correlated common factors, so that

$$(5.2)\quad \Sigma=A\Psi A'+U^2, \text{ with } \mathrm{var}(\xi)=\Psi\neq I_m.$$

In this case Ψ can be factored again for a smaller set of "second order factors" which then, presumably, have the status of even more fundamental variables than the "primary factors" x_k in ξ.

In addition, I believe it is necessary to mention two further aspects of Thurstone's work which also left a longlasting mark on the future course of factor theory, and psychometrics in general:

(v) Thurstone's scientific perspective could be described, perhaps, as solipsistic. In taking credit for the m-variate generalization of Spearman's TFT ("...I discovered that the tetrad was merely the expansion of a second order minor, and the relation was then obvious...If the second order minors must vanish in order to establish a single common factor, then must the third order minors vanish in order to establish two common factors, and so on?" (1947, p. iv)), he did not mention that much the same thing had already been shown by Garnett in (1919) and had been elaborated upon in Dodd's (1928) review: "The generalized criterion for n variables to be factorable into n-m factors is that every multiple correlation ratio or coefficient of the (n-m)th order shall equal unity. In the case of rectilinear regression, the more convenient criterion is that every discriminant of the (n-m)th order shall equal zero" (Dodd, 1928, p.275).

Similarly, in discussing various "regression methods for estimating factor scores" in (1938, Ch. 10; and 1947, Ch. 21) Thurstone does not mention Wilson's work on factor indeterminacy, which might cast doubts on the logical justification of such "estimation", nor Piaggio's (1931) or Thomson's (1934) work on the same topic, which had covered the same algebra from a more sophisticated point of view: "Spearman's case is exactly the same [as the regression case] except for the important fact that he has no criterion, no measure of 'g' except through the team of tests themselves...This distinction between the two cases may appear to be subtle, but it seems a proper distinction to draw" (Thomson, 1934, p. 94f; my emphasis).

As a result, many earlier contributions to factor theory were eventually all but forgotten. They have only recently been retrieved again (see, e.g., Schönemann, 1971; Steiger and Schönemann, 1978), against considerable resistance, because in the meantime a great deal of effort and personal prestige had been invested into pursuits which have led nowhere. This waste could have beeen avoided with a more openminded attitude towards Thurstone's predecessors.

(vi) Thurstone, on occasion, used unconventional language which was apt to create the impression of scientific achievements for which he had little or no empirical evidence.

Problem (v) was probably aggravated by Wolfle's (1940) "Factor Analysis to 1940", to which McNemar alludes in the above quote. Although this historical account "provides the specialist in factor analysis with a single list of publications of the 1928-1940 period...only the more important or the more pertinent ones are referred to in the review" (p.5). Among the casualties of this eclectic treatment were all discussions of factor indeterminacy, which had appeared during this very review period, together with most other contributions of a more critical nature by Wilson, Camp, Spearman, Mackie and others. Since it was "the purpose of this review to survey the literature that has accumulated since Dodd's 1928 review" (p.1), a more appropriate title might have been "Factor analysis from 1929 to 1940". The lack of specificity in the title may have contributed to the subsequent neglect of Dodd's excellent and far more complete review of the earlier history. Neither Thurstone (1947) nor Harman (1967) mention it, though both refer to Wolfle's.

An illustration of (vi) is Thurstone's ambiguous use of the term "law", as in his "Law of Comparative Judgement". He stated it in a paper with this title (1927a) as a "law", whose "derivation will not be repeated here because it has been described in a previous article" (p. 41). But in the previous article (Thurstone, 1927b), this "law" is only a "reasonable...assumption subject to verification in every case" (p. 24). There is no additional empirical evidence in either article to justify the semantic transition from a "reasonable assumption" to a "new psychophysical law" (1927a, p.39). Nor has there been much progress since on the question how this "law" can be "verified", or at least differentially falsified against entirely different "laws", e.g. Luce's Choice Theory, which usually explains the same data equally well. However, in the pursuit of this difficult question Burke and Zinnes (1965) found that some of Thurstone's (1927c) data fit neither "law".

A charitable interpretation of this loose choice of words might be that Thurstone meant by "law" not a replicable empirical relationship of some generality (as, in, e.g. "Weber's Law"), but rather a normative law, a guiding principle. The trouble with such semantic laxity is that it makes it difficult to know whether a stated result is still in need of empirical confirmation or not. At least this proved difficult for many of Thurstone's disciples.

An important example of this is Thurstone's celebrated Simple Structure principle, which virtually all of his followers adopted with little if any attempt to test its substantive validity. Numerous programs were written in the late 50's and early 60's which simply imposed this "law" on the data. Since Thurstone intended the Simple Structure principle as a device for defining the factors as intersections of hyperplanes, it might seem reasonable that they first would would have checked whether there were indeed well-defined hyperplanes in the data.

There is nothing intuitively obvious about the expectation that all test vectors should align themselves in linear subspaces, e.g. in 3 intersecting planes, if there are three common factors. If one shoots a burst of insecticide with a spraycan into the air one would be rather surprised if the droplets were to arrange themselves in 3 planes. It is more likely that they will uniformly fill out a roughly conical part of the space. This expectation is the baseline against which Bargmann (1955) tests the actually achieved simple structure: he stretches each testvector in the common factor space to unit length and then checks how many of them are contained in the circular disks defined by the m subsets of m-1 factors. On the hypothesis of a uniform distribution of the extended testvectors on the surface of the unit sphere, one can compute the probability that r out of p test vectors will be contained in a disk of an arbitrarily chosen width. Before claiming Simple Structure, one should at least make sure that there are "significantly more" testvectors in each disk than can be expected by chance alone.

The significance of Bargmann's contribution eluded most factorists who relied on the simple structure principle for defining their factors. An exception was R. B. Cattell: "Nevertheless, by any standards of true scientific procedure an SS significance test is an essential part of a factor analytic experiment" (Cattell, 1978, p. 176). However, on using it, Cattell soon discovered that "the Bargmann test in practice has turned out to be a severe critic, and it is no exaggeration to say that about half of the investigators publishing studies would have their conclusions wiped out thereby – which is perhaps why a test of significance of the obtained simple structure is so rarely published" (p. 175). For example, on applying Bargmann's test to Thurstone's celebrated Primary Mental Ability study (1938), one will find that none of the thirteen hyperplanes reach significance at the .05 level. Although the situation had much improved in the 1941 replication of this study by Thurstone and Thurstone (all but one of the ten hyperplanes reach significance), the reduced study for 21 selected tests again gives only 4 out of 8 significant hyperplanes by Bargmann's criterion. The point is not, of course, that sometimes our hypotheses turn out to be false, but rather that the post-Thurstonian generation of factorists has acted as if their hypotheses always had to be true.

Very much the same uncritical attitude prevails, to this day, in regard to the second hypothesis implied by Thurstone's multiple factor generalization of Spearman's TFT: How are we to decide how many common factors are in the data? And can it ever happen that the CFM simply does not fit, regardless how many factors are postulated? The answer to the second question is, unfortunately, "no". This is in contrast to Spearman's model, which, at least in principle, was either true or false. Thurstone's "generalization" can only be conditionally falsified, relative to a prespecified number of common factors m_0.

Prior to the development of the presently popular iterative algorithms for fitting and testing the multiple factor model, one simply extracted as many common factors as seemed necessary to give the residual matrix a pleasing appearance. Usually this required about a third as many factors as observed variables.

More recently, it has become customary "to guess" m_0 in advance, and then to test the hypothesis H_0: $m=m_0$ statistically after m_0 common factors have been extracted. Before examining the merits of these statistical tests more closely in the next section, I would like to discuss briefly one of the presently popular devices for guessing the correct number of common factors, which is sometimes affectionately called "Kaiser's Little Jiffy":

"It is proposed, then, that m, the number of factors to be retained in the varimax method be given by the number of latent roots of the correlation matrix under consideration which are larger than one...This rule of behavior for 'when to stop factoring' is purely logical – there is, for example, no readily apparent mathematical rationale for choosing it. It does, however, seem to reflect the sort of thinking which is implicit traditionally in factor analysis" (Kaiser, 1956, p. 10). In 1974, Kaiser and Rice thought they had discovered a more convincing rationale for this rule in Guttman (1954): "The writers adopt as the answer to the crucially important question of the 'number of factors' Guttman's (1954) classic lower bound, the index of the covariance matrix (with zeroes in the diagonal) under consideration. This answer is the same as that given by Kaiser's (1956, 1960, 1970) extensively used 'eigenvalues greater than one of R'" (p. 111).

This attractively simple answer to the crucially important number of factors question has indeed been extensively used. Maxwell (1972), for example, tried this "popular rule, attributed to Guttman (1954), for interpreting the results of a principal component analysis [which] is to retain only those components for which the latent roots are greater than unity" (p. 13). Many computer packages have incorporated this simple rule as a stock answer for the crucial question how many factors to retain in the CFM.

While it cannot be ruled out that this popular rule may occasionally give the correct answer (see Appendix A4 for some data on its performance when the number of factors is known), I doubt that Guttman has ever recommended it or supplied the mathematical rationale for it. Since we are fortunate enough to have him sitting right here, we can, of course, ask him about this.

One reason why this popular rule on deciding on the number of common factors is apt to be unreliable is that it does not make much sense statistically:

Consider an N-fold sample from a population with covariance matrix $\Sigma=I_p$. The sample correlation matrix R will be different from the identity matrix with probability one for any finite sample size N. Since the sum of the latent roots equals the trace of R, which is p, their average must be unity and, since $R\neq I$, some roots must be larger and some must be smaller than one. Hence, the number of common factors according to Kaiser's rule is at least one. This is hardly a "lower bound" for the true number of common factors, because the true number of factors is zero in this case.

The reason for this discrepancy between Guttman's theory and Kaiser's and Maxwell's practice is that in (1954) Guttman was not interested in the sample case: "In the present paper, we do not treat the problem of ordinary sampling error, that is, of sampling a population of respondents. We assume throughout that population parameters are used, and not sample statistics" (p. 151).

Far from wanting to contribute to the growing mechanization of factor analytic research by means of this or that Little Jiffy, Guttman was actually trying to sound a warning against the unreflecting acceptance of Thurstone's parsimony principle as the sole basis for the post-Thurstonian phase of "exploratory factor analysis": "The question as to whether a parsimoneous common-factor system exists at all for a given set of data remains a fundamental one to be reexplored in each empirical case. Current computing procedure which aim at stopping at some relatively small number of common factors may prejudge the issue" (p. 160). Although this was written 25 years ago, this is precisely what most presently popular computer packages encourage the user to do.

6. JöRESKOG: ANSWERS ONLY HALF THE QUESTION

In this section I wish to return once more to the "crucially important number of factors question". This question is indeed important if one takes the factor model at all seriously as a mathematical theory for psychology: "If we grant that men are not all equal in intellectual endowment and in temperament and if we have the faith that this domain can be investigated as a science, then we must make the plausible and inevitable assumption that individual differences can be conceived in terms of a finite number of traits, parameters or factors" (Thurstone, 1947, p.58). However, as Guttman (1954) has observed, "psychologists cannot invoke general algebra nor can algebraists invoke psychology to make out an a priori case for small rank" (p. 302), because "algebra and psychology both indicate large rank to be the more proper null hypothesis problem for mental test data" (p. 307). Another reason why we should be more careful in the future with making "plausible and inevitable assumptions" without serious attempts to test them has been stated by Russell: "The method of 'postulating' what one wants has many advantages; they are the same advantages of theft over honest toil" (1919, p. 71). Since psychology in general, and the IQ industry in particular, have of late increasingly come under public scrutiny we may not much longer be able to afford to act as if we did not know this difference.

There are other, more technical reasons for being concerned about thecrucial number of factors question if one subscribes to the Spearman/Thurstone interpretation of the factor model. For example, if the true number of factors is overestimated, then the factor pattern becomes unestimable statistically, because then it is no longer uniquely defined (Anderson and Rubin, 1956). Similarly, the success of Williams' attempt to eliminate "...the so-called problem of factor score indeterminacy" (1980, p. 303) hinges critically on the number of factors question, because he "will always assume that for every p it is possible to write

(6.1) $V_p = F_p F_p' + U_p U_p'$

<u>where F_p is a pxr real matrix</u>" (p. 294, my emphasis. His V stands for my Σ. The subscript denotes the number of tests in a given study). How are we to test this assumption in practice which implies, among other things, that the number of factors, r, (my m), does not change with p?

On paper the answer to this question is deceptively simple. It has been worked out a long time ago by Lawley (1940). However, few have bothered to ask how good this answer actually is. Since the CFM (3.1), (3.2) implies that the partial correlation matrix

(6.2) $\mathrm{corr}(\eta|\xi) = U^{-1}(\Sigma - A\,A')U^{-1} = I_p$,

is diagonal, this necessary condition can, at least in principle, be tested as a statistical hypothesis in applications of this model if one appends the usual distribution assumptions of multinormality (e.g. Lawley, 1940; Howe, 1955; Bargmann, 1957).

Sophisticated and expensive computer algorithms for performing factor analyses within such a presumably more rational statistical framework have been widely used in the social sciences for the last 15 years. Bentler (1980) gives an up-to-date summary of these developments.

These programs provide the user with a chi-square test of fit which permits control over Type I Error. However, if one is at all concerned about the social consequences faulty psychological models have had in the past (e.g., Kamin, 1974), and may have in the future (e.g., Jensen, 1969), then it would seem rather obvious that Type II Error, the error of promulgating an incorrect theory, is the more important problem. Surprisingly, however, after the promising beginnings made by Browne (1968) and Linn (1968), studies which critically evaluate the power of these tests are hard to find. To worry about Type I Error when testing models, just because this error happens to be easier to control, is like searching for a lost dime under a lamppost, just because the light is better there.

It was therefore a welcome change of pace when Geweke and Singleton recently (1980) published the results of a Monte Carlo study designed to evaluate the performance of the likelihood ratio test (LRT) of unrestricted maximum likelihood factor analysis (UMLFA) in small samples. They generated two population covariance matrices according to the CFM (3.3) (with Ψ=I) for the factor patterns

(6.3) $A_1' = (4\ 4\ 3\ 3\ 2)$, $A_2' = \begin{pmatrix} 4 & 4 & 3 & 3 & 2 \\ 1 & 2 & 2 & 1 & 2 \end{pmatrix}$

and the matrix of unique variances

(6.4) $\mathrm{diag}\,(U^2) = \mathrm{diag}\,(2.0,\ 1.66,\ 1.33,\ .66,\ .33)$.

On drawing random samples of various sizes and factoranalyzing the resulting sample covariance matrices, the authors arrived at the following overall conclusion:

"The likelihood ratio test has considerable power even when sample size is only 10. In the experiment reported here, the first factor explained 70 percent of the variance of the indicator, on the average, while the second factor explained 22 percent of the variance; yet even with a sample of only 10 observations the one factor model was rejected at the 5 percent level in 24 percent of the replications" (p. 136).

The concern of these authors for the important and somewhat neglected power problem in UMLFA is a positive step which, hopefully, will be followed up with similar studies in the future, to enable users of such programs to evaluate their merit on more solid grounds than blind faith (see Steiger and Lind, 1980, for some recent work on this problem). However, future studies should try to avoid a flaw which impairs the practical relevance of the Geweke and Singleton study: their generating unique variances (6.4) imply for the factor pattern A_2 communalities which are very unlikely to arise in real life: the average communality is .93.

A rough estimate of the communality range encountered in the social sciences can be gleaned from a review of some 69 factor analytic studies conducted in the forties which has been compiled by French (1951). The distribution of the average communalities for the 61 studies with admissible communality estimates is given in Appendix A5.1.

In no case does the average communality reach the value employed in the Geweke and Singleton study. The two highest values stay well below .75, and the average is around .55.

To obtain an idea what the power of the LRT might be for the factor pattern A_2 in (6.3) for communalities in the more typically encountered range, random samples were drawn from multinormal populations with covariance matrices

(6.5) $\Sigma_k = A_2A_2' + q_k U^2$,

for various values of the multiplier q_k, which controls the effective average communality in the population. 200 such samples were reanalyzed for $q_k=1$ (i.e. in exact accord with the population matrix selected by Geweke and Singleton) under H_0: m = 2, to estimate the null distribution of the test statistic for A_2, which has rank 2. The rest of the simulation dealt with the (incorrect) hypothesis

(6.6) H_0: m = 1,

to assess the power of the test as a function of the average communality for fixed A_2. In all cases but one the sample size was fixed at N=30. The number of replications for each q_k was 200. The UMLFA algorithm was written by Browne (1969).

The results of this replication of the Geweke and Singleton study for varied communality parameters are portrayed in Appendix A5.2 in terms of the cumulative proportions of the probabilities associated with the chi-square statistics obtained on reanalysis of the sampled correlation matrices. The data points are based on 100 replications each, to convey an impression of the stability of such empirical estimates for 100 replications, the number employed in the cited study.

Although the power estimates have not yet fully stabilized, the trend is sufficiently clear to warrant an overall conclusion which is at variance with that drawn by Geweke and Singleton. The left hand side of the graph A5.2 gives the power estimates for their population covariance matrix for N=30. Geweke and Singleton found that "for 30 observations the...proportions [of rejecting H_0 at the .05 and .01 level] were 63 and 42 percent, and for 150 and 300 observations the one factor model was always rejected at both significance levels". Superficially, the present reanalysis seems to corroborate their finding, because at the .1 level the power was estimated as .80 and .79 for each of the 2 runs with 100 replications.

However, as the graph shows, this impressive result is tied to the unrealistic choice of the particular population covariance matrix by Geweke and Singleton. As can be seen from the graph, the power fades rapidly as the average communality approaches a more realistic range. When this value attains .72, which is still at the upper tail of the distribution for the French summary, the power estimates for

.1 level tests are in the upper 10's or low 20's. For the modal value of the French distribution (h^2=.55), the power barely exceeds the chosen level of significance. This disappointing performance of the test of fit improves only slightly when the sample size is raised from 30 to 100. At this point a .2 level test has a chance of 1 out of 3 to detect that H_0: m = 1 is false. A theoretical reason for this intimate relation between the power of the LRT in UMLFA and the average communality is given in (Schönemann, 1980).

Although these empirical results are conditional on the particular factor pattern A_2 selected by Geweke and Singleton, they nevertheless raise sufficient doubts about the value of the LRT in UMLFA in general to call for further, more representative investigations of the actual statistical merits of maximum likelihood algorithms for factor analysis, which, so far, seem to have been largely taken for granted.

Thus, although psychologists have studied the factor model for more than 50 years, and statisticians have joined them for at least 25 years, we still do not know how to test the basic assumption this model is built on with any stringency. To be precise: we now have expensive computer programs for evaluating the probability of falsely rejecting a model we would like to accept, but we still have no idea how often it has been falsely accepted in the past.

7. JENSEN: CRYSTALLIZED ANSWERS TO FLUID QUESTIONS

"For the first time, intelligence testing has a firm foundation, and for the first time there is a genuinely operational definition for intelligence. The assessment of the intellectual resources of man can now take on features of psychoengineering...Talents of numerous kinds can be discovered in individuals, and their development can be promoted because their properties are known. Optimal development through education and optimal placement of the individual within the scheme of things can now be more nearly achieved, and these steps should contribute to the satisfaction for all concerned" (Guilford and Hoepfner, 1971, p. 361).

Between Guilford and Hoepfner's echo of Hart and Spearman's prophesies in (1912; see section 2) lies half a century of intelligence research. "Few topics have been as avidly researched, at all levels of sophistication" (Wechsler, 1971, p. 50). During this time, Wechsler's test "has been administered to millions of patients by clinicians around the world, in many languages, and in many countries on both sides of the iron curtain" (Edwards, 1974, p. 429).

During the last two decades or so we have witnessed a growing sense of disillusionment both with most extant intelligence theories (e.g., Tuddenham, 1962; Hudson, 1972) and with the results of the post-Thurstonian period of "blind factor analysis" (Mulaik, 1972, p. 9; Mulaik, 1978) in general. In his delightful personal account Hudson summarized his frustations this way: "It can scarcely be coincidental that psychologists who have measured children's intelligence have armoured themselves to a greater extent than any other with the protective magic of number. Nor can it be coincidental that in the course of half a century, the mental test movement has told us little about children that we did not already know, but has made a major contribution in the field of statistics" (1972, p.55).

In the preceding pages I have tried to show that the method of factor analysis, which Spearman proposed at the beginning of this period for operationally defining "intelligence", has been misunderstood and misused by many researchers in the intelligence field. I have further tried to show that this method does not achieve its designated objective of providing an operational definition of "intelligence", and that this method is beset with numerous problems of its own which only now are beginning to surface again, after they had been ignored for several decades.

Recent events have also drawn public attention to the intelligence field which can no longer be ignored (Kamin, 1974; Hearnshaw, 1979; Nairn, 1980). These events are unrelated to the technical deficiencies I have discussed here. In particular, Jensen has stirred up widespread controversy with the views he expressed in his (1969) Harvard Educational Review article "How much can we boost IQ and scholastic achievement?". He has drawn criticism both from outside and from within psychology: "Arthur Jensen speaks to many of us - I hope I am not among them, but who can be sure?" (Hudson, 1972, p. 123).

I believe some of this criticism may be unfairly directed at Jensen. Although his conclusions may not be representative of his field, his methods for arriving at them are. Not being a specialist in methodology, Jensen has to rely, as have many others before him, on research tools others have put at his disposal, and on methodological myths others have generated for him. According to one modern text, "the formulation of the factor model approaches that of quantum theory" (Rummel, 1970, p.28). For Rummel "factor analysis is a general scientific method for analyzing data" (p.13). These are almost exactly Thurstone's words (1947, p. 55). The uncritical acceptance of such myths is widespread, because they are convenient and enobling. They are convenient because they promise a mechanical device for conducting research. They are enobling because they lend the mantle of scientific legitimacy to unproven conjectures.

The technical jargon of the method of analysis can be used to imbue isolated facts of questionable reliability and dubious relevance with added significance. Take, for example, the following factual observation by Eysenck:

"Eskimos, living in the icy wastes far above the arctic circle, score at or above white Canadian norms on the Progressive Matrices" (1973, p. 484). There is little dispute over this and similar facts. The dispute revolves around what they mean. In and of itself the fact that Eskimos score the same as some other group on a particular test is of no interest or consequence whatsoever. It is not clear, for example, what possible benefit the Eskimos derive from their performance on this test, other than a citation by Professor Eysenck. Any additional benefit would have to be established through painstaking experimental research, on Eskimos. Since, as Jensen (1980, p.314) informs us, the Matrices correlate only .55 with the Wechsler Bellevue, and in some instances .53 with the WAIS, and .27 with the Stanford Binet, the Eskimos' performance on the Matrices would not even enable us to predict their performance on the other IQ tests very well, if that ever were of interest. Therefore, before it makes sense to ask how much we can boost the IQ of Eskimos, we would first have to explain which of the four IQs we wish to boost, and then why.

If, on the other hand, we wave the magic wand of factoranalytic jargon, then such simple facts can be made to sound quite imposing:

"Factorially the Progressive Matrices apparently measures g and little else...Many other tests also measure g to a high degree, but few if any, have so little loadings as the Raven on any of the main group factors - Verbal, Numerical, Spatial and Memory...Some of the small spurious factors that emerge from factor analysis of the inter-item correlations are not really ability factors at all but are 'difficulty factors', due to varying degrees of restriction of variance of items on widely differing difficulty levels and to nonlinear regressions of item difficulties on age and ability (McDonald, 1965). When proper account is taken of these psychometric artifacts, the RPM seems to measure only a single factor of mental ability, which is best called g. The raw inter-item correlation matrix on large unselected samples closely approximates the appearance of a "simplex", which means that all the intercorrelations can be 'explained' most parsimoneously in terms of a single factor plus random errors of measurement" (Jensen, 1980, p. 646f).

Most laymen (and not a few psychologists) will stand aghast at so much technical language. They will walk away impressed with the high state of the art of intelligence research it seems to reflect.

Few readers will notice, for example, that Guttman (1958) had actually presented "The simplex as a counterexample" (p. 303) to Thurstone's hypothesis "that relatively small rank could be attained for correlation matrices of mental test data by use of communalitites" (p. 297). Few will be familiar with Thomson's and Mackie's reservations about the psychological implications of correlation matrices which actually do have small rank. And few will know that the "crucial" question of how to decide how many factors there are in a given set of data is still wide open, despite the best efforts of a number of professional statisticians, because they have had only time to answer the wrong half of the question, so far.

However, as the record shows, there has been a longstanding tradition in psychometrics to eliminate all disquieting discussions of facts and interpretations which conflict with Thurstone's parsimoneous theories. It is, therefore, not surprising that most users of factor analysis are unaware of its limitations.

ACKNOWLEDGEMENTS

I would like to dedicate this paper to Louis Guttman, who has been a lonely beacon of sanity in a sea of confusion for many years.

I would like to thank my wife, Roberta, for her unstinting support in hard times and for her help in editing the manuscript. I also would like to express my appreciation to the students in my 1980 factor analysis class for help with the literature search, the collection and analysis of some of the data used here, and for their enthusiasm and indulgence with my idiosyncracies (many of them are clinical students). Finally, I would like to thank Professor H. Feger for including me in this illustrious gathering, and my good friend Ingwer Borg for having brought me to his attention.

REFERENCES

Anderson, T.W. and Rubin, H. Statistical inference in factor analysis. Proceedings of the 3rd Berkeley Symposium on Mathematical Statistics and Probability, 1956, 5, 111.

Bargmann, R. Signifikanz-Untersuchungen der einfachen Struktur in der Faktorenanalyse. Mitteilungsblatt fur Mathematische Statistik. Wurzburg: Physica Verlag, 1955.

Bargmann, R. A study of independence and dependence in multivariate normal analysis. University of North Carolina, Institute of Statistics Mimeo Series No. 186, 1957.

Bartlett, M.S. The statistical conception of mental factors. Journal of Psychology, 1936, 28, 97-103.

Bentler, P.M. Multivariate analysis with latent variables: Causal modeling. Annual Review of Psychology, 1980, 31.

BJMSP, Editorial communication. British Journal of Mathematical and Statistical Psychology, May 22, 1978.

Blum, J.M. Pseudoscience and Mental Ability. New York and London: Monthly Review Press, 1978.

Boring, E.G. Intelligence as the tests test it. New Republic, 1923, 35, 35-37.

Browne, M.W. Fitting the factor analysis model. Psychometrika, 1969, 34, 375-394.

Browne, M.W. A comparison of factor analytic techniques. Psychometrika, 1968, 33, 267-334.

Burke, C.J. and Zinnes, J.L. A paired comparison of pair comparisons. Journal of Mathematical Psychology, 1965, 2, 53-76.

Cattell, R.B. Factor analysis. New York: Harper, 1952.

Cattell, R.B. The scientific use of factor analysis in behavioral and life sciences. New York and London: Plenum Press, 1978.

Cronbach, L.J. Essentials of psychological testing. New York: Harper and Brothers, 1949.

Dodd, S.C. The theory of factors. Psychological Review, 1928, 35, 211-234, 261-279.

Edwards, A.J. Introduction. In: Selected papers of David Wechsler. New York, London: Academic Press, 1974.

Eysenck, H.J. The measurement of intelligence. Baltimore: Williams and Wilkins Company, 1973.

French J.W., The description of aptitude and achievement tests in terms of rotated factors. Psychometric Monographs, No.5, 1951.

Fruchter, B. Introduction to factor analysis. New York: Van Nostrand, 1954. (Revised, 1964).

Garnett, J.C.M. General ability, cleverness, and purpose. British Journal of Psychology, 1919, 9, 345-366.

Garnett, J.C.M. The single general factor in dissimilar mental measurements. British Journal of Psychology, 1920, 10, 242-258.

Geweke, J.F. and Singleton, K.J. Interpreting the likelihood ratio statistic in factor models when sample size is small. Journal of the American Statistical Association, 1980, 75, 133-137.

Guilford, J.P. and Hoepfner, R. The analysis of intelligence. New York: McGraw-Hill, 1971.

Guttman, L. Multiple-group methods for common-factor analysis. Psychometrika, 1952, 17, 209-222.

Guttman, L. To what extent can communalities reduce rank? Psychometrika, 1958, 23, 297-308.

Guttman, L. Some necessary conditions for common-factor analysis. Psychometrika, 1954, 19, 149-161.

Guttman, L. The determinacy of factor score matrices with implications for five other basic problems of common-factor theory. British Journal of Statistical Psychology, 1955, 8, 65-81.

Guttman, L. The Green-McDonald proof of the nonexistence of factor analysis. Unpublished manuscript, 1975.

Guttman, L. What is not what in theory construction. Paper read at the Eighth Annual Meeting of the Israel Sociological Association, Haifa, February, 1977.

Harman, H.H. Modern factor analysis. Chicago: University of Chicago Press, 1960. (Revised, 1967).

Hart, B. and Spearman, C. General ability: Its existence and nature. British Journal of Psychology, 1912, 5, 51-84.

Hearnshaw, L.S. Cyril Burt, psychologist. Ithica: Cornell University Press, 1979.

Howe, W.G. Some contributions to factor analysis. Report No. ORNL-1919, Oak Ridge National Laboratory, Oak Ridge, Tennessee, 1955.

Hudson, L. The cult of the fact. New York, Evanston, San Francisco: Harper and Row, 1972.

Hunsaker, J. and MacLane, S. Edwin Bidwell Wilson. Biographical Memoirs in National Academy of Sciences, 1973, 43, 284-320.

Jensen, A.R. How much can we boost IQ and scholastic achievement? Harvard Educational Review, 1969, 39, 1-123.

Jensen, A.R. Bias in mental testing. New York: Free Press, 1980.

Kaiser, H.F. The varimax method of factor analysis. Ph.D. Thesis, University of California, 1956.

Kaiser, H.F. The application of electronic computers to factor analysis. Educational and Psychological Measurement, 1960, 20, 141-151.

Kaiser, H.F. A second generation Little Jiffy. Psychometrika, 1970, 35, 401-415.

Kaiser, H.F. and Rice, J. Little jiffy. Mark IV. Journal of Educational and Psychological Measurement, 1974, 34, 111-117.

Kamin, L.J. The science and politics of I.Q. Potomac, Maryland: Lawrence Erlbaum Associates, 1974.

Kruskal, J.B. A mathematical discussion of McDonald's Second Draft (Item 25). In: Final Report by the Ad Hoc Committee to Consider Certain Charges to the Editorial Council of the Psychometric Society, 1978.

Lawley, D.N. The estimation of factor loadings by the method of maximum likelihood. Proceedings of the Royal Society of Edinburgh, Section A, 1940, 60, 64-82.

Lawley, D.N. and Maxwell, A.E. Factor analysis as a statistical method. London: Butterworths, 1971.

Linn, R.L. A Monte Carlo approach to the number of factors problem. Psychometrika, 1968, 33, 37-71.

Mackie, J. The sampling theory as a variant of the two-factor theory. Journal of Educational Psychology, 1928, 19, 614-621.

Maxwell, A.E. Factor analysis: Thomson's sampling theory recalled. British Journal of Mathematical and Statistical Psychology, 1972, 25, 1-21.

McDonald, R.P. Difficulty factors and nonlinear factor analysis. British Journal of Mathematical and Statisical Psychology, 1965, 18, 11-23.

McDonald, R.P. Some uncertainties about factor indeterminacy. Unpublished manuscript, 1972.

McDonald, R.P. The measurement of factor indeterminacy. Psychometrika, 1974, 39, 203-221.

McNemar, Q. The factors in factoring behavior. Psychometrika, 1951, 16, 353-359.

Mulaik, S.A. The foundations of factor analysis. New York: McGraw-Hill, 1972.

Mulaik, S.A. Traits, semantics and psychometrics. Paper read to the Society for Multivariate Experimental Pscyhology. Cincinnati, Ohio, November 1978.

Mulaik, S.A. and McDonald, R.P. The effect of additional variables on factor indeterminacy with a single common factor. Psychometrika, 1978, 43, 177-192.

Nairn, A. and Associates. The reign of ETS. The Ralph Nader Report on the Educational Testing Service, 1980.

Piaggio, H.T.H. The general factor in Spearman's theory of intelligence. Nature, 1931, 127, 56-57.

Rummel, R.J. Applied factor analysis. Evanston: Northwestern University Press, 1970.

Russell, B.R. Introduction to mathematical philosophy. London: George Allen and Unwin LTD, 1919.

Schönemann, P.H. The minimum average correlation between equivalent sets of uncorrelated factors. Psychometrika, 1971, 36, 21-30. Erratum. Psychometrika, 1973, 38, 149.

Schönemann, P.H. Some nontrivial implications of factor indeterminacy. Paper read at the Annual Meeting of the Society for Multivariate Experimental Psychology. Cincinnati, Ohio, June 1978.

Schönemann, P.H. A note on power as a function of communality in factor analysis. Submitted to Journal of the American Statistical Association, April, 1980.

Schönemann, P.H. and Wang, M.M. Some new results on factor indeterminacy. Psychometrika, 1972, 37, 61-91.

Schönemannm P.H. and Steiger, J.H. Regression component analysis. British Journal of Mathematical and Statistical Psychology, 1976, 29, 175-189.

Schönemann, P.H. and Steiger, J.H. On the validity of indeterminate factor scores. Bulletin of the Psychonomic Society, 1978, 12, 287-290.

Spearman, C. General intelligence, objectively determined and measured, American Journal of Psychology, 1904, 15, 201-293.

Spearman, C. The abilities of man. New York: Macmillan Company, 1927.

Steiger, J.H. The relationship between external variables and common factors. Psychometrika, 1979, 44, 193-197.

Steiger, J.H. and Schönemann, P.H. A history of factor indeterminacy. In S. Shye (ed.): Theory construction and data analysis in the behavioral sciences. San Francisco: Jossey Bass, 136-178, 1978.

Steiger, J.C. and Lind, J.M. Statistically-based tests for the number of common factors. Paper read at the Annual Spring Meeting of the Psychometric Society, Iowa City, 1980.

Thomson, G.H. A hierarchy without a general factor. British Journal of Psychology, 1916, 8, 271-281.

Thomson, G.H. The general factor fallacy in psychology, British Journal of Psychology, 1920, 10, 319-326.

Thomson, G.H. The meaning of "i" in the estimate of "g". British Journal of Psycholgy, 1934, 25, 92-99.

Thomson, G.H. The factorial analysis of human ability. (5th ed.) Boston: Houghton Mifflin, 1951.

Thorndike, E.L. Educational psychology, New York: Teachers College, Columbia University, 1903.

Thurstone, L.L. A law of comparative judgment. Psychological Review, 1927a, 34, 273-286. In: Thurstone, L.L. The measurement of values. Chicago: University of Chicago Press, 1959.

Thurstone, L.L. Psychophysical analysis. American Journal of Psychology, 1927b, 38, 368-389. In: Thurstone, L.L. The measurement of values. Chicago: University of

Chicago Press, 1959.

Thurstone, L.L. The method of paired comparisons for social values. Journal of Abnormal and Social Psychology, 1927c, 21, 384-400. In: Thurstone, L.L. The measurement of values. Chicago: University of Chicago Press, 1959.

Thurstone, L.L. The vectors of mind. Chicago: University of Chicago Press, 1935.

Thurstone, L.L. Primary mental abilities. Chicago: University of Chicago Press, 1938.

Thurstone, L.L. Multiple-factor analysis. Chicago: University of Chicago Press, 1947.

Thurstone, L.L. The measurement of values. Chicago: University of Chicago Press, 1959.

Thurstone, L.L. and Thurstone, T.G. Factorial studies of intelligence. Chicago: University of Chicago Press, 1941.

Tuddenham, R.D. The nature and measurement of intelligence. In L. Postman (Ed.), Psychology in the making: Histories of selected research problems. New York: Knopf, 1962.

Vernon, P. Intelligence. In: Dockrell, W.B. (ed.) The Toronto Symposium on Intelligence. Toronto: Ontario Institute for Studies in Education, 1969.

Vernon, P.E. Intelligence: Heredity and environment. San Francisco: W.H. Freeman and Company, 1979.

Wechsler, D.W. Manual for the Wechsler Adult Intelligence Scale, New York: Psychological Corporation, 1955.

Wechsler, D.W. Manual for the Wechsler Intelligence Scale for Children - revised. New York: Psychological Corporation, 1974b.

Wechsler, D. Rationale of the children's scale. What intelligence tests test. In: Selected papers of David Wechsler. New York, London: Academic Press, 1974a.

Wechsler, D. The measurement of adult intelligence, 1st ed. Baltimore: Williams and Wilkins, 1939. In Selected papers of David Wechsler. New York, London: Academic Press, 1974.

Wechsler, D. The measurement and appraisal of adult intelligence. (4th ed.) Baltimore: Williams and Wilkens,

1958.

Williams, J.S. A definition for the common-factor analysis model and the elimination of problems of factor score indeterminacy. Psychometrika, 1978, 43, 293-306.

Wilson, E.B. The so-called foundations of geometry. Archiv der Mathematik und Physik, 1903, 6(3), 104-122.

Wilson, E.B. Probable inference, the law of succession and statistical inference. Journal of the American Statistical Association, 1927, 22, 209-212.

Wilson, E.B. Review of "The abilities of man, their nature and measurement", by C. Spearman. Science, 1928, 67, 244-248.

Wolfle, D. Factor analysis to 1940. Psychometric Monographs, No. 3, 1940.

APPENDIX A1: THOMSON'S SAMPLING THEORY

A1.1: EXAMPLE OF A SAMPLING PATTERN S (eq. 3.5):

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	P
1		1		1	1	1	1		1	1	1	1	1		1	1	1	1	1	1	.8
2	1	1	1		1	1	1	1		1		1	1	1	1		1	1	1	1	.8
3	1	1	1	1	1		1	1	1		1	1	1	1	1	1	1	1		1	.8
4	1				1	1		1	1				1		1			1		1	.4
5	1							1					1	1	1			1			.4
6								1		1								1	1		.4
7		1	1																		.2
8							1	1		1	1							1		1	.2
9						1		1					1			1		1		1	.2

A1.2: POPULATION CORRELATION MATRIX (eq. 3.9) above diagonal. ASSOCIATED SAMPLE CORRELATION MATRIX (N = 200) below diagonal.

	1	2	3	4	5	6	7	8	9
1	1.00	.750	.788	.583	.306	.375	.177	.510	.510
2	.716	1.00	.788	.667	.612	.500	.354	.510	.510
3	.765	.769	1.	.647	.594	.243	.343	.495	.495
4	.616	.675	.705	1.	.680	.333	.000	.408	.680
5	.283	.595	.593	.714	1.	.480	.000	.333	.500
6	.369	.491	.205	.360	.381	1.	.000	.612	.408
7	.133	.313	.265	.022	-.047	-.021	1.	.000	.000
8	.519	.524	.432	.465	.311	.663	-.073	1.	.500
9	.548	.489	.501	.689	.516	.452	-.027	.570	1.

A1.3: SUMMARY OF MONTE CARLO STUDY FOR VARIOUS VALUES OF m (N = 200):

m/m_o:	1	2	3	4	5
20	.39				
	.85				
	.31				
	.00	.00	.00*	.00*	.00*
100	.00	.01	.08*	.22*	
200	.94				
	.52				
	.02	.21*	.34*		
	.00	.22	.46		

m/m_o:	1	2	3	4	5
1000	.33				
	.39				
	.85				
	.31				
	.14	.68			
	.00	.01	08*	22*	
	.00	00	.00*	.00*	.00*
2000	.45				
	.64				
	.71				
	.60				
	.70				

Entries are probabilities of UMLFA LRT for $H_0: m = m_o$. *boundary cases

APPENDIX A2: FACTOR ANALYSIS AND REGRESSION COMPONENT ANALYSIS OF THE WAIS

A2.1 CORRELATIONS (Source: Wechsler, 1974b, Table 9). 150 males and 150 females, ages 45-54. Above diagonal.
REPRODUCED CORRELATIONS MLE $\hat{\Sigma} = A A' + U^2$ under H_0: m = 2; below diagonal.

	1	2	3	4	5	6	7	8	9	10	11
1	.174	.740	.710	.720	.550	.850	.580	.650	.550	.600	.540
2	.745	.327	.640	.660	.490	.760	.510	.630	.490	.580	.510
3	.688	.627	.395	.580	.600	.680	.540	.570	.580	.530	.500
4	.715	.647	.609	.376	.450	.720	.570	.620	.500	.570	.490
5	.546	.499	.490	.487	.601	.530	.480	.510	.420	.430	.450
6	.846	.760	.691	.726	.543	.126	.630	.630	.520	.640	.500
7	.612	.559	.546	.545	.444	.611	.505	.550	.500	.590	.480
8	.654	.602	.603	.590	.496	.644	.551	.376	.590	.630	.570
9	.544	.510	.548	.510	.463	.514	.510	.603	.361	.520	.640
10	.631	.577	.567	.563	.462	.628	.514	.574	.536	.466	.500
11	.529	.495	.530	.495	.448	.500	.494	.582	.615	.518	.408

COMMUNALITIES:

1	2	3	4	5	6	7	8	9	10	11
.826	.673	.605	.624	.399	.874	.495	.624	.639	.534	.592

The MLE is used as a population correlation matrix for the score simulation in Appendix A3.

APPENDIX A2.2: FACTOR PATTERN, FACTOR STRUCTURE AND REFERENCE VECTOR STRUCTURE AFTER OBLIQUE (PROMAX) ROTATION TO SIMPLE STRUCTURE:

		FACTOR PATTERN A:		FACTOR STRUCTURE G		REFERENCE VECTOR STRUCTURE	
		V	P	V	P	V	P
1	Information	.841	.088	.907	.723	.551	.057
2	Comprehension	.713	.135	.815	.674	.467	.089
3	Arithmetic	.485	.344	.745	.710	.318	.226
4	Similarities	.638	.188	.780	.670	.418	.123
5	Digit Span	.317	.357	.587	.596	.208	.234
6	Vocabulary	.961	-.035	.934	.691	.630	-.023
7	Digit Symbol	.378	.373	.660	.658	.248	.244
8	Pict. Completion	.272	.564	.698	.770	.178	.370
9	Block Design	-.090	.865	.564	.797	-.059	.567
10	Pict Arrangement	.367	.413	.679	.690	.240	.271
11	Object Assembly	-.072	.822	.549	.768	-.047	.539

APPENDIX A2.3: REGRESSION COMPONENT ANALYSIS OF THE WAIS DATA IN A2.1:

RCA (Schönemann and Steiger, 1976) is a data reduction method patterned after Guttman's (1952) Multiple Group Method. It closely parallels factor analysis, except that the new variables ("regression components") are linear combinations of the observed variables y_i, and hence are determinate. The defining equations of RCA are (i) $\eta = A\xi + \varepsilon$, (ii) $\xi = B'\eta$, $A = \mathrm{cov}(\eta,\xi)\ \mathrm{var}^{-1}(\xi)$. The analysis can either be based on a set of "defining weights" B (which then determines A), or on a "regression pattern" A (which then determines B). In the present case it is based on A, obtained on rotating the regression pattern for the first standardized Principal Components obliquely to Simple Structure.

	REGR. PATTERN A		DEFINING WEIGHTS B	
	V	P	V	P
1	.913	.001	.211	-.045
2	.878	-.027	.204	-.055
3	.636	.221	136	.063
4	.824	.013	.190	-.035
5	.390	.343	073	.127
6	.981	-.080	231	-.083
7	.502	.296	101	.101
8	.365	.520	.059	.204
9	-.071	921	-.062	.396
10	.478	.348	.093	.125
11	.116	.950	.074	.411

r_{vp} = .698

Although the numerical results usually are very similar to those obtained with the CFM, the theoretical interpretation is quite different: in particular, there is no claim that such regression components have any scientific status of their own (as factors are supposed to), because RCA is purely tautological, not inferential. Any scientific significance of the components would have to be established through independent experimental work.

APPENDIX A3: FACTOR INDETERMINACY

A3.1: FIRST SET OF EXACT FACTOR SCORES ($X_1 Z$) for 20 SUBJECTS WHICH REPRODUCE THE MLE IN A1.2 EXACTLY:

Subj:	X_1	X_2	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8	Z_9	Z_{10}	Z_{11}
1	-.428	.138	-1.202	-.668	-.401	.144	1.432	.436	.765	-1.665	.503	1.868	-.742
2	-.842	.035	.751	-1.051	.758	1.098	-1.643	.607	1.272	.982	.728	.007	-.357
3	.483	.647	-.246	.228	-1.523	.092	.915	.046	.914	.369	-.448	1.311	1.835
4	1.002	.168	1.096	.846	1.857	1.439	1.983	-.650	.874	-.301	-.758	-.419	-.734
5	.326	-.435	-.674	.154	-1.046	-.727	-.680	-1.402	-.232	1.299	.370	-.011	.462
6	-.029	1.080	.008	.776	.368	-1.380	.717	-.957	-.794	.072	1.718	-1.284	-1.562
7	-1.791	-2.239	-.517	-1.144	-.213	-1.516	.101	-.619	1.811	.151	0.448	-.990	-.661
8	.876	1.213	.022	-.154	.714	-1.787	-.637	2.257	.070	-.514	-2.007	-.760	.122
9	-.192	.329	.263	-1.075	.971	.517	-.118	-1.649	-1.302	1.178	-.905	.591	.826
10	.002	.673	2.138	.447	-1.207	-.181	-.333	-1.047	.202	-1.593	-.395	-1.134	1.451
11	-.235	.132	.993	-.147	-.749	1.403	-.173	1.083	1.014	1.098	.701	-.188	-.462
12	.367	.081	-.763	-.791	-.159	1.079	-1.224	.004	-.927	-2.297	1.398	-.410	.214
13	2.618	1.689	-1.022	-.302	-.656	.649	-1.201	-.296	-.022	.372	-.788	-.196	-1.432
14	.585	.476	-.244	.929	2.204	-1.315	-.995	-.409	.791	-.070	1.027	1.754	1.034
15	.111	.258	-.566	-2.356	.315	-.065	1.526	-.037	-1.003	.207	-.352	-.589	.525
16	-1.011	-2.179	.035	.291	.906	.503	-.150	1.121	-1.105	-.330	.413	-.667	1.179
17	-1.529	-.666	-.129	1.107	-.246	.331	-.049	.800	-1.971	.633	-.912	1.001	-.803
18	.878	.365	-.145	.838	-.731	-.389	1.192	1.452	-.238	1.259	1.572	-.714	.676
19	-.089	-1.081	2.013	.024	-.934	-.859	-.273	-.107	-.693	-.494	-.247	1.600	-1.541
20	-1.102	-.685	-1.812	2.050	-.228	.964	-.389	-.633	.573	-.357	-1.170	-.771	-.030

APPENDIX A3.1 (continued): SECOND SET OF (MINIMALLY CORRELATED) EXACT FACTOR SCORES (X^*, Z^*) FOR 20 SUBJECTS WHICH REPRODUCE THE MLE IN A1.2 EXACTLY:

Subj.:	X^*_1	X^*_2	Z^*_1	Z^*_2	Z^*_3	Z^*_4	Z^*_5	Z^*_6	Z^*_7	Z^*_8	Z^*_9	Z^*_{10}	Z^*_{11}
1	-.393	-.172	-1.207	-.638	-.258	.202	1.560	.311	.909	-1.395	.955	2.037	-.339
2	-.176	.121	-.610	-1.900	.197	.379	-1.955	-1.185	.873	.608	.704	-.403	-.392
3	.523	1.198	-.443	.047	-1.856	-.118	.645	-.008	.603	-.156	-1.236	.955	1.130
4	1.386	.671	.214	.247	1.284	.885	1.594	-1.641	.406	-.934	-1.424	-.930	-1.337
5	-.576	.250	1.002	1.116	-.725	.002	-.626	1.106	-.112	1.068	-.751	.060	-.522
6	-.204	.398	.503	1.155	.876	-.989	1.103	-.552	-.343	.777	2.674	-.778	-.704
7	-2.348	-2.171	.590	-.467	.179	-.958	.297	.892	2.071	.335	-.629	-.732	-.812
8	1.363	-.163	-.669	-.435	1.092	-1.871	-.203	.806	.534	.536	.046	-.189	1.947
9	-.506	.343	.894	-.687	1.206	.840	.004	-.797	-1.142	1.304	-.972	.752	.772
10	.084	.059	2.102	.490	-.934	-.078	-.084	-1.328	.480	-1.064	.500	-.807	2.250
11	.567	.379	-.675	-1.204	-1.503	.493	-.615	-1.060	.458	.515	.465	-.768	-.690
12	-.154	-.088	.323	-.101	.336	1.673	-.933	1.396	-.561	-1.909	1.564	-.027	.373
13	1.919	.970	.538	.739	.277	1.596	-.585	1.523	.727	1.343	.142	.614	-.587
14	.570	1.613	-.453	.679	1.593	-1.648	-1.513	-.256	.202	-1.110	-.613	1.074	-.431
15	-.327	.092	.353	-1.770	.744	.442	1.782	1.132	-.683	.554	-.178	-.253	.689
16	-.821	-1.055	-.585	-.212	.144	-.040	-.745	.717	-1.796	-1.448	-1.176	-1.449	-.245
17	-.868	-1.470	-1.292	.474	-.315	-.110	.051	-1.065	-1.900	1.080	.344	1.132	.307
18	1.145	1.442	-.908	.251	-1.526	-.996	.587	.839	-.944	.150	.062	-1.508	-.679
19	.009	-1.154	1.832	-.081	-.969	-.938	-.279	-.378	-.707	-.469	-.127	1.592	-1.435
20	-1.191	-1.264	-1.512	2.297	.157	1.233	-.086	-.452	.924	.215	-.350	-.373	.705

APPENDIX A3.2: SIMULATED EXACT TESTSCORES Y WHICH REPRODUCE THE MLE IN A1.2 EXACTLY AS $\hat{\Sigma} = YY/N$. Note that $Y = AX+UZ = AX^*+UZ^*$ FOR A, U AS IN A2.2 AND (X, Z), (X*,Z*) IN A3.1

	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}	y_{11}
1	-.849	-.669	-.412	-.159	1.024	-.261	.434	-1.059	.460	1.175	-.329
2	-.392	-1.197	.080	.143	-1.528	-.594	.599	.393	.543	-.290	-.138
3	.360	.562	-.500	.486	1.094	.457	1.073	.723	.247	1.339	1.670
4	1.314	1.221	1.710	1.554	1.915	.726	1.063	.183	-.399	.151	-.403
5	-.045	.262	-.649	-.319	-.579	-.169	-.204	.639	-.183	-.067	-.086
6	.073	.570	.589	-.662	.932	-.406	-.173	.645	1.968	-.442	-.108
7	-1.918	-2.235	-1.773	-2.494	-1.289	-1.863	-.225	-1.658	-2.045	-2.258	-2.134
8	.853	.701	1.291	-.308	.217	1.602	.834	.608	-.235	.304	1.012
9	-.023	-.708	.630	.256	-.035	-.782	-.875	.855	-.242	.469	.812
10	.952	.348	-.526	.017	-.017	-.393	.395	-.596	.344	-.496	1.480
11	.228	-.234	-.539	.735	-.162	.155	.681	.684	.556	-.160	-.170
12	-.003	-.180	.106	.911	-.804	.351	-.490	-1.262	.877	-.111	.177
13	1.923	1.922	1.438	2.386	.501	2.351	1.604	1.893	.753	1.524	.284
14	.432	1.013	1.832	-.344	-.416	.400	.961	.385	.976	1.609	1.010
15	-.120	-1.234	.340	.079	1.311	.085	-.575	.303	.302	-.255	.540
16	-1.027	-.850	-.671	-.747	-1.214	-.497	-1.980	-1.707	-1.546	-1.726	-.965
17	-1.398	-.547	-1.125	-.898	-.761	-1.161	-2.227	-.403	-.987	-.153	-.950
18	.710	1.155	.092	.391	1.333	1.347	.299	1.217	1.182	-.014	.669
19	.670	-.196	-1.002	-.787	-.625	-.085	-.929	-.937	-1.075	.613	-1.867
20	-1.742	.295	-.913	-.241	-.895	-1.260	-.264	-.905	-1.197	-1.213	-.503

APPENDIX A3.3: EXACT COMMON FACTOR SCORES (X), MINIMALLY CORRELATED EQUIVALENT COMMON FACTOR SCORES (X*), "FACTOR SCORE ESTIMATES" ($\hat{X}$), AND REGRESSION COMPONENT SCORES CORRESPONDING TO WAIS DATA IN A2.1 AFTER CONVERSION INTO IQ'S (MEAN 100, STANDARD DEVIATION 15). EQUIVALENT IQ'S WHICH DIFFER BY MORE THAN 1/2 STANDARD DEVIATION SINGLY UNDERLINED, THOSE DIFFERING BY MORE THAN 1 STANDARD DEVIATION DOUBLY UNDERLINED:

	x_1	x_2	x_1^*	x_2^*	$\hat{x}_1$	$\hat{x}_2$	x_1^c	x_2^c
1	93.578	102.075	94.101	97.424	93.840	99.750	95.500	103.420
2	87.371	100.532	97.358	101.818	92.365	101.175	92.425	103.015
3	107.239	109.703	107.840	117.977	107.540	113.840	108.070	118.210
4	115.024	102.526	120.796	110.061	117.910	106.294	123.235	99.265
5	104.896	93.469	91.361	103.752	98.128	98.611	97.630	98.380
6	99.563	116.193	96.943	105.965	98.253	111.079	98.890	114.580
7	73.128	66.412	64.783	67.435	68.955	66.923	67.795	67.660
8	113.146	118.197	120.439	97.561	116.793	107.879	113.710	107.200
9	97.122	104.932	92.405	105.142	94.764	105.037	96.430	107.680
10	100.033	110.093	101.257	100.890	100.645	105.491	99.100	108.055
11	96.480	101.975	108.502	105.688	102.491	103.831	102.455	103.735
12	105.501	101.219	97.687	98.674	101.594	99.947	99.565	99.295
13	139.264	125.332	128.781	114.554	134.023	119.943	135.625	112.585
14	108.773	107.141	108.546	124.199	108.660	115.670	110.185	117.175
15	101.671	103.872	95.094	101.380	98.383	102.626	96.940	106.675
16	84.830	67.314	87.685	84.170	86.257	75.742	83.170	72.835
17	77.069	90.014	86.978	77.948	82.024	83.981	82.210	84.160
18	113.177	105.482	117.169	121.632	115.173	113.557	112.885	114.625
19	98.668	83.788	100.134	82.686	99.401	83.237	97.945	77.065
20	83.466	89.730	82.142	81.044	82.804	85.387	86.260	84.385

APPENDIX 3.4: FACTOR INDETERMINACY INDICES FOR THE WISC-REVISED*:

SOURCE: WECHSLER, 1974c

Age	m_o	(orthogonal) VARIMAX solution smallest minimum correlation	No. of discrepancies exceeding 1/2 st. dev.	(oblique) PROMAX solution smallest minimum correlation	No. of discrepancies exceeding 1/2 st. dev.
11.5	4	.15	16/20	.42	14/20
12.5	3	.56	8/20	.55	10/20
13.5	3	.26	12/20	.71	12/20
14.5	4	.08	14/20		
	3	.49	13/20		
	2	.54	13/20		
15.5	4	.28	18/20	.62	9/20
	3	.59	12/20	.73	11/20
16.5	4	.63	9/20		
	3	.28	9/20		
	2	.59	12/20	.74	7/20

Note that oblique rotation, in general, improves the definition of the common factors for the same m(see Schönemann and Wang, 1972, p. 70) as well as the definition of Simple Structure (not shown). Extracting more common factors usually depresses the indeterminacy measures, except when boundary cases arise (the frequency of which increases with m_o) and the corresponding factor is partialled out as a component. The author's reservation about such a mixture of factor and component analysis are stated in loc. cit., p. 87.

*The author is indebted to Mr. Gary Hu for some of these computations.

APPENDIX A 4: IN A JIFFY

24 correlation matrices were generated according to the CFM with $A = (a_{ij})$ drawn from a uniform distribution (over the (0,1) interval) for various values of m and p =10. U^2 was scaled in each case so that the average population communality was exactly .6. Two samples each were then generated for the sample sizes N = 100, 200, 400, 800, and the eigenvalues were computed. There was no systematic sample size effect. The contingency table gives the frequencies of the number of eigenvalues exceeding unity (KLJ m) for three values of true (ideal) rank m.

	TRUE m:				
KLJ m	2	4	6		
1	3	3	5	11	
2	5*	4	3	12	
3	0	1	0	1	
	8	8	8	24	*: correct decision

APPENDIX A5: POWER AS A FUNCTION OF COMMUNALITY IN UMLFA

A5.1: AVERAGE COMMUNALITIES IN 61 STUDIES REVIEWED BY FRENCH (1951)

Upper boundary:	.35	.40	.45	.50	.55	.60	.65	.70	.75
Class Frequency:	02	04	03	12	19	10	07	02	02

COMMUNALITIES IMPLIED BY VARIOUS VALUES OF q_k FOR $\Sigma_k = A_2 A_2' + q_k U^2$;
FOR THE GEWEKE AND SINGLETON FACTOR PATTERN A2;

q_k	1	2	3	4	5	$\overline{h^2}$	
1.0	.90	.92	.91	.94	.96	.93	(Geweke and Singleton)
1.5	.85	.89	.87	.91	.94	.89	
2.0	.81	.86	.83	.88	.92	.86	
3.0	.74	.80	.77	.84	.89	.81	
5.0	.63	.71	.66	.75	.83	.72	
7.0	.55	.63	.58	.68	.78	.65	
10.	.46	.55	.49	.60	.71	.56	(French survey)

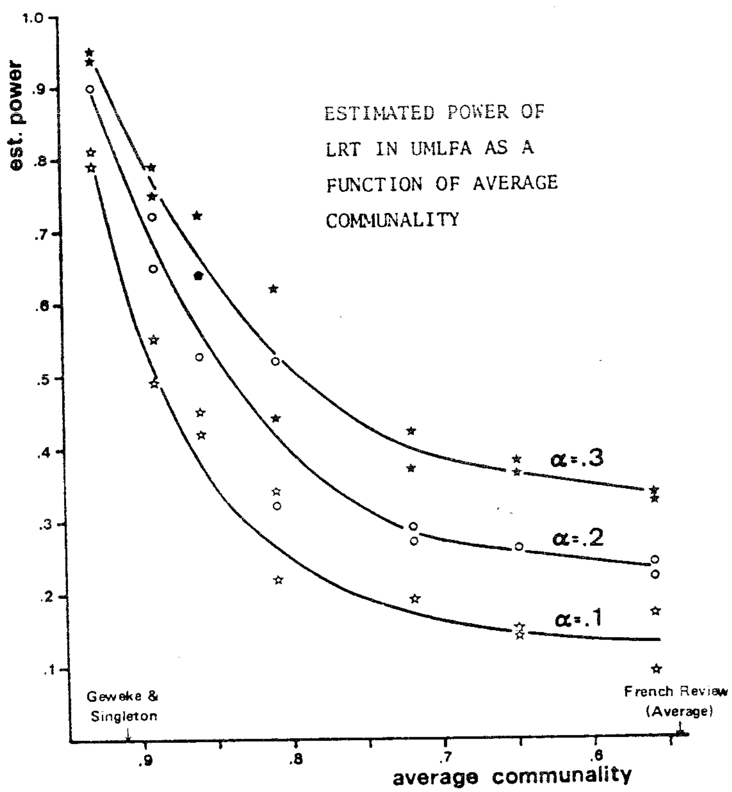

Figure 1

Graph shows percent of rejections of H_0: m=1 against H_1: m=2 (true) for factor pattern A_2 (eq. 6.3) in Geweke & Singleton study (1980) as a function of average communality (see Table A.5). Power is given for 3 alpha levels. Each data point is based on 100 replications, so that curves are interpolated averages for 200 replications.

Note that power is high only for unrealistically high average communality (left side of graph), and falls off sharply as average communality approaches more realistic values. For the average of the French survey (1951), estimated power exceeds barely the alpha level of the test (right side of graph).

Reprinted by permission from: *Science*, 1928, *LXVII*, 244-248.

Edwin Bidwell Wilson was born in 1879 and died in 1964. "Wilson was the last student of J. Willard Gibbs at Yale and had worked creatively in many fields of mathematics and physics; his advanced calculus was a standard text for decades; he wrote one of the earliest texts on aerodynamics; he was a friend of R. A. Fisher and an expert on mathematical statistics and demography. My early formulation of the inequality in Eq. 4 owed much to Wilson's lectures on thermodynamics" [Paul A. Samualson, Nobel address, reprinted in *Science*, 1971, *173*, 993-994].

If asked to list the four most important pioneers of factor analysis, I would name Spearman, Thomson, Wilson and Guttman. Of the four, only Spearman kept his faith in the model, the others became more and more disenchanted as they learned more about it. Today, only three are reasonably well-known. At the time, Spearman and Thurstone recognized Wilson's stature at once: "The development of factor theory as well as its applications in science, will be accelerated by the assistance of mathematicians; and it is gratifying that Professor E. B. Wilson has turned his attention to these problems in several papers" [Thurstone, *Vectors of the Mind*, 1935]. This enthusiasm waned quickly when Wilson pointed out several flaws which had escaped notice of the factorists of the preceding decades.

Only a dozen of Wilson's more than 200 papers deal with factor analysis. In those few, he raised all the major problems which still face us today. Until I rediscovered Wilson in the stacks of the Ohio State Library in the late 60's, his name had virtually disappeared from all standard texts on factor analysis. Wilson was never given the recognition he deserved by the Psychometric community, which, perhaps, is only logical. — *P. H. S.*

SCIENTIFIC BOOKS

The Abilities of Man, their Nature and Measurement. By C. Spearman. New York, The Macmillan Co., 1927. vi + 415 + xxxii pp.

The Grote professor of philosophy of mind at the University of London has written an important book. It could not be otherwise when the book represents the cumulation of intellectual endeavor for a period of a quarter century by such as he. It may well be that he does not know exactly what his theories and facts signify; it is certain that I do not. The work has been supported during its progress by the collaboration of a multitude of Spearman's pupils and by others, it has drawn widely upon the investigations of other schools, it has also had constant opposition and the book has been severely criticized in a review in *Nature* (August 6, 1927, p. 180) which has led to an interchange of views between author and reviewer (*Nature*, November 12, 1927, p. 690). Into this difference I will not enter except to say that whether the book is mathematically complete or not does not interest me; this is unimportant. Science advances not so much by the completeness or elegance of its mathematics as by the significance of its facts. You can not upset the findings of the "Origin of Species" either by the contraposition of your religious convictions or by observing that Darwin's statistical technique was not up to standard. Science goes forward upon "evidence beyond reasonable doubt"; to that evidence incomplete mathematics may contribute valuable elements.

Spearman's chief thesis is that when a group of persons x, y, z, . . . are given a test *a*, say of arithmetic or spelling or literary interpretation, the marks m_{ax}, m_{ay}, m_{az}, . . . which they score represent in part their respective general intelligences g_x, g_y, g_z, . . . and in part their special abilities in the subject, s_{ax}, s_{ay}, s_{az}, . . . This would seem incontrovertible provided we mean by ability in the subject, ability to get scores in the test. The necessity for this proviso may be illustrated as follows. I have some general intelligence; I have some mathematical ability; yet if an examiner should set me a mathematical test in Yiddish, which might be "easy meat" for a lot of candidates for admission to our colleges, I should miserably fail. It may further be remarked that the scores m_{ax}, m_{ay}, . . . may depend on the manner of scoring used by the examiner or his clerk. For example, if the test be of the simple sort where a large number of questions are answered yes or no, one method of scoring is to count the number of right answers, R_x, R_y, . . . ; another method is to take the difference between the numbers right and wrong $(R-W)_x$, $(R-W)_y$, ... If all the N questions are answered, the scores are equivalent since $W = N - R$ and the series of scores R_x, R_y, . . . and $2R_x - N$, $2R_y - N$, . . . are in the same order, will give the same correlations with other tests, etc. But if some of the questions are unanswered (U), the second series becomes $2R_x - U_x - N$, $2R_y - U_y - N$, . . . which need not be equivalent to R_x, R_y, ... How are we to compare the answers of two persons to 50 questions if one answers 40 all correctly and the other answers all 50 with 45 right and 5 wrong?

The next thesis is that when a battery of tests *a*, *b*, . . . are sufficiently different, so that the scores may be assumed to have in common only the general intelligence we may write for the nk marks of the n individual x, y, z, . . . on the k tests *a*, *b*, . . .

$$m_{ax} = c_a g_x + s'_{ax} \qquad m_{bx} = c_b g_x + s'_{bx}$$
$$m_{ay} = c_a g_y + s'_{ay} \qquad m_{by} = c_b g_y + s'_{by} \qquad (1)$$

in such a manner that the general intelligence g and the special abilities s′, are uncorrelated, *i.e.*,

$$\Sigma g_x s'_{ax} = 0, \qquad \Sigma g_x s'_{bx} = 0, \ldots \qquad (2)$$
$$\Sigma s'_{ax} s'_{bx} = 0, \qquad \Sigma s'_{ax} s'_{cx} = 0, \ldots \qquad (3)$$

when the summation runs over the individuals x, y, z, ... This leads to some correlation algebra to prove both that such a resolution of the marks is possible and that it is unique. I have read the proofs with care (including the references to the literature, not all of which has been reproduced in the book) and have found no errors in the mathematics. Yet I am not entirely happy, satisfied. I should like to have found at least one example worked out in detail—one set of nk scores for n individuals on k tests worked through to the determination of the n values $g_x, g_y, \ldots$ of the general intelligences of those individuals and of the nk values $s'_{ax}, s'_{ay}, \ldots; s'_{bx}, s'_{by}, \ldots; \ldots$ of their special abilities on each of the tests. Theorems which prove the existence of some possibility do not satisfy the practical applied mathematician—we do not so much want to know that there is a solution to the problem as to know what the solution is! I will work an example below.

What solution does the author offer us? (First he adopts scales which reduce the scores on each test so that they have the same dispersion about their means, we may take it as unity, which is also the dispersion of g.) If r_{ag} be the correlation coefficient between g and the test a, he shows that the solution is

$$g_x = r_{ag} m_{ax}$$

with a probable error of .6745 $(1 - r^2_{ag})^{1/2}$. Note that the answer is a regression equation. We do not know the individual values $g_x, g_y, \ldots$; we could write

$$g_x = r_{ag} m_{ax} + e_{ax}$$
$$g_y = r_{ag} m_{ay} + e_{ay}$$

where $e^2_{ax} + e^2_{ay} + \ldots = n(1 - r^2_{ax})$. If the author desires to prove that testing does not determine the general intelligence of the persons tested he has succeeded. Why did he pick on test *a* to determine g? Evidently one could equally well write

$$g_x = r_{bg} m_{bx}, \text{ etc.}$$

with a probable error .6745 $(1 - r^2_{bg})^{1/2}$, etc. Practically we might choose that test as *a* which has the highest correlation with g. Better, he shows how to weight the different tests so as to get a combined score *t* which best determines g. In the example this best score gives $r_{tg} = .75$ so that the probable error in g_x is $.6745 \times .6614 = .446$. When we recall that the scale of g is such that the standard deviation of g is unity or that one half of the n values of g lie between $-.67$ and $+.67$, two thirds of them between -1.0 and $+1.0$, we can appreciate the significance of a probable error of .45. The solution for the special ability is likewise

$$s_{ax} = (1 - r^2_{ag})^{1/2} m_{ax}$$

with the probable error .6745 r_{ag}. The better the test *a* estimates g the worse it estimates the special ability. Spearman's comment is: "We are faced by the fact that the current measurements of specific abilities—upon which have come to hang the weal or woe of countless individuals in industry and otherwise—are little more than the blind leading the blind." Rather pessimistic I call it, possibly unjustifiably so in view of such success as persons like O'Connor (West Lynn Works, General Electric Company) have in their placement work.

Spearman gives a long discussion of the attempts that have been made to define general intelligence. He does not define it, he computes it, and at that only by a regression equation, he does not measure it any more than he would weigh a person by computing his weight from his height through a regression equation of weight on height. He sets **forth** a hypothesis that the general intelligence is energy, the special abilities are engines, with apparently the will as engineer. This is allegory. If intelligence were energy it should be measured in ergs—but again he calls it a force (p. 414), so perhaps he thinks of measuring it in dynes. Or perchance the whole is mere logomachy. It would be interesting to enquire which of the technical physical terms is most like g, the general intelligence. Perhaps it might be efficiency! It would also be interesting to know just what he or Maxwell Garnett (a competent applied mathematician) means by the word unique in the proof that the resolution into g's and s's is unique. He can hardly mean that the regression equation $g_x : \sigma_g = r_{ag} m_{ax} : \sigma_a$ is unique since there is one such for each test and they give different results. If he means that given the nk grades $m_{ax}, m_{bx}, \ldots, m_{ay} \ldots$ we can determine the actual values of $g_x, g_y, \ldots$, why are we given the regression?

Example (preamble). If we can assign the k quantities $c_a, c_b, \ldots$ and the n values $g_x, g_y, \ldots$ equations (1) will determine the nk special abilities s' from the nk grades m. Equations (2) if the n values $g_x, g_y, \ldots$ are known will determine the k values $c_a, c_b, \ldots$ as $c_a = \sigma_a r_{ag} / \sigma_g$. To have the intelligence g on a uniform scale we shall assume $\sigma_g = 1$ which gives one quadratic equation between the n values $g_x, g_y, \ldots$

$$g^2_x + g^2_y + g^2_z + \ldots + g^2_n = n \qquad (4)$$

We must find the k coefficients r_{ag}. Equations (3) when expressed in terms of the m's and g's give the equations

$$r_{ag} r_{bg} = r_{ab}, \qquad r_{ag} r_{cg} = r_{ac} \qquad (3)$$

and if there be three or more tests enable us to solve for r_{ag}, etc., as

$$r_{ag} = \sqrt{\frac{r_{ab} r_{ac}}{r_{bc}}} \text{etc.} \qquad (5)$$

This requires that the values r_{ag} should be fractional or (if all the correlations $r_{ab}, r_{ac}, r_{bc}, \ldots$ between tests be positive, as is usually the case) that

$$r_{bc} \gtreqless r_{ab} r_{ac} \qquad (6)$$

or that the partial coefficients $r_{bc.a}$ shall all be positive, and, if there are more than three tests, that the so-called tetrad relations vanish, *i.e.*,

$$r_{ac} r_{bd} - r_{ad} r_{bc} = 0. \qquad (7)$$

These relations (6) and (7) are verified within the experimental error with respect to a large variety of intelligence tests. There is the equation

$$g_x + g_y + g_z \ldots + g_n = 0 \qquad (8)$$

introduced to simplify the analysis and refer all g's to their mean. If the m's are also thus expressed, as is most convenient, the s's will be relative to their means. The k equations (5) are linear in the g's, *viz.*,

$$m_{ax} g_x + \ldots + m_{an} g_n = n\, r_{ag} \sigma_a \qquad (5')$$

We have, therefore, in (5′), (8) and (4) the number $k+1$ of linear equations and one quadratic equation in the n quantities g. It would seem as though the n values g could be found with $n-k-2$ degrees of freedom, *i.e.*, that, as n is generally much larger than $k+2$, the solution should be indeterminate rather than unique.

Example (solution). Try a case. Let the marks of 6 students on 3 tests be (the first columns give the actual marks, the second columns give the differences from the means)

	a		b		c	
1	10	5	8	3	1	2
2	8	3	5	0	9	4
3	6	1	9	4	4	−1
4	4	−1	7	2	8	3
5	2	−3	0	−5	1	−4
6	0	−5	1	−4	1	−4
$6\sigma^2$		70		70		62

$$r_{bc} = .66, \quad r_{ac} = .73 \quad r_{ab} = .74$$

The equations to be solved are

$$5g_1 + 3g_2 + g_3 - g_4 - 3g_5 - 5g_6 = 18.5$$
$$3g_1 + 0g_2 + 4g_3 + 2g_4 - 5g_5 - 4g_6 = 16.7$$
$$2g_1 + 4g_2 - g_3 + 3g_4 - 4g_5 - 4g_6 = 16.3$$
$$g_1 + g_2 + g_3 + g_4 + g_5 + g_6 = 0$$
$$g^2_1 + g^2_2 + g^2_3 + g^2_4 + g^2_5 + g^2_6 = 6$$

The result of solving the first four for g_1, g_2, g_5, g_6 in terms of g_3 and g_4 and substituting in the last is

$$g_3 = .49 - .7g_4 \pm .89_4 \sqrt{(-g^2_4 + .40g_4 + .069)}$$

The radical is positive only if g_4 lies between $-.14$ and $+.52$ and any value of g_4 between these limits is possible. For the two limits the solutions for the g's are

		diff.
$g_1 = .91$	$g_1 = 1.43$	−.5
$g_2 = 1.21$	$g_2 = .42$	+.8
$g_3 = .59$	$g_3 = .12$	+.5
$g_4 = -.14$	$g_4 = .52$	−.7
$g_5 = -1.47$	$g_5 = -.80$	−.7
$g_6 = -1.10$	$g_6 = -1.69$	+.6

Notice that the ranges of possible intelligence for the six are different; we have a better line on 1 and 3 than on 6 and know least about 2.

Let us next compute the special abilities so standardized that their standard deviations are unity. The equations given by Spearman are like

$$m_{a1}/\sigma_a = r_{ag} g_1 + \sqrt{1 - r^2_{ag}}\, s_{a1}$$
$$m_{a1}/3.4 = .905\, g_1 + .42\, s_{a1}$$
$$s_{a1} = .7\, m_{a1} - 2.1\, g_1$$

On the basis of the extreme alternative solutions given above we have

		diff.
$s_{a1} = 1.6$	$s_{a1} = .4$	+1.2
$s_{a2} = -.4$	$s_{a2} = 1.2$	−1.6
$s_{a3} = -.5$	$s_{a3} = .5$	−1.0
$s_{a4} = -.4$	$s_{a4} = -1.8$	+1.4
$s_{a5} = +1.0$	$s_{a5} = -.4$	+1.4
$s_{a6} = -1.2$	$s_{a6} = +.1$	−1.3

Similarly we could compute for tests *b* and *c* the limits of specific ability. (The calculations given above have been carried to so few places that a positive check can not be expected, either for the zero mean or the unit standard deviation.) What we have shown is that the complete solution can be obtained but is indeterminate. We have had no need of any harder mathematics than the solution of a set of k+1 linear equations and 1 quadratic equation. We do not need the generalized Bravais distribution (as used by Garnett) and in view of Yule's wise comments on mental measurements (*Brit. J. Psych.*, vol. 12, p. 100.) to all of which I hereby subscribe, it would seem quite superfluous to introduce this higher mathematics, involving a probability theory which probably does not apply anyhow, to make determinate (if it does) that which without it seems indeterminate.

Do $g_x, g_y, \ldots$ whether determined or undetermin able represent the intelligence of x, y, . . . ? The author advances a deal of argument and of statistics to show that they do. This is for psychologists, not for me, to assess. I believe he does not adequately emphasize the fact that they represent the intelli gence only relative to the set-up of the tests. That this is so is evident from general considerations of the transformation theory of correlation algebra; but as even the term "transformation theory of correlation algebra" seemed unknown and unintelligible to a large group of persons professionally interested in statistics and in education when I recently mentioned it to them, I take the space, in a review already too long, to expound the obvious. 1°, If we have nk marks of n individuals x, y, z, . . . on k tests *a*, *b*, *c*, . . . we may combine these marks into new sets of scores, call them *a′*, *b′*, *c′*, . . . , in a linear fashion as

$$m'_{ax} = c_{11}\, m_{ax} + c_{12} m_{bx} + \dots$$
$$m'_{ag} = c_{11}\, m_{ay} + c_{12} m_{by} + \dots$$

$$m'_{bx} = c_{21}\, m_{ax} + c_{22} m_{by} + \dots$$

$$m'_{cx} = c_{31}\, m_{ax} + c_{32} m_{by} + \dots$$

with k^2 constants c_{ij}. These new scores m′ contain all the information of the old scores because they can be solved for the old scores m, but the information is differently assembled. It may be that these scores do not measure any particular ability such as spelling or literary interpretation or mathematical judgment, but they do represent scores involving certain weighted combinations of such abilities and with my limited knowledge of intelligence testing seem to represent some sorts of ability. 2°, Irrespective of whether the tetrad relations (7) are fulfilled, we can choose the constants c_{ij} in infinitely many ways so that the new scores are all uncorrelated, *i.e.*, $r'_{ab} = r'_{ac} = r'_{bc} = \dots = 0$. In this case the tetrad equations for the new correlation coefficients must vanish. Now if $g_x, g_y, \dots$ be the general intelligence of the persons tested the equations (3′) can no longer be solved as in (5) for $r'_{ag}, r'_{bg}, \dots$ because (5) become indeterminate; but inspection of the equations

$$r'_{ag}\, r'_{bg} = r_{ab} = 0$$
$$r'_{ag}\, r'_{cg} = r_{ac} = 0$$

shows that all of the correlations of g to the new scores must vanish except at most one. Which one? As the equations defining *a′*, *b′*, *c′*, . . . are largely arbitrary, the symmetrical and natural conclusion would be that none of them are correlated with g. Or we might so form one of them say *a′* as to agree that it represents the intelligence g with $r'_{ag} = 1$ and the others represent special abilities independent of g. Next, 3°, to be more specific we may take as one simple definition of *a′*, *b′*, *c′*, . . .

$$m'_{ax} = m_{ax}$$
$$m'_{bx} = m_{bx} - \sigma_b r_{ab} m_{ax}/\sigma_a$$
$$m'_{cx} = m_{cx} + \beta m_{bx} + \alpha m_a$$

and determine β, α, so that $r'_{ac} = 0$, $r'_{bc} = 0$, and so on. Now as we know r_{ag} by (5) as other than 0, it follows that $r'_{ag} = r_{ag} \neq 0$ and that the remaining values r'_{bg}, r'_{cg}, . . . all vanish. But from the definition of *b′*

$$0 = \sigma'_b\, r'_{bg} = \sigma_b (r_{bg} - r_{ab} r_{ag})$$

or

$$r_{bg} - r_{ab} r_{ag} = 0$$

This last equation is, however, impossible since we know that $r_{ag} r_{bg} = r_{ab}$. Hence, 4°, any set of values $g_x, g_y, \dots$ for the general intelligence of x, y, . . . which will go with the set-up of tests *a*, *b*, *c*, . . . can not possibly go with the set-up *a′*, *b′*, *c′*, . . . but must be replaced by new values $g'_x, g'_y, \dots$ approximate to that set-up. Yet the information we had in the nk scores of x, y, . . . on *a*, *b*, . . . is all contained in the nk scores assigned to x, y, . . . on *a′*, *b′*, . . .; the persons x, y, . . . are the same but their intelligences have changed—the old values whether indeterminate or unique will no longer do. What does this leave of the concept of the intelligence of an individual x as measured by g_x? Apparently only that it is relative to the set-up, which is the obvious proposition that I set out to prove.

The intelligence tester may object that the scores on *a′*, *b′* . . . mean nothing, are mere artificialities, whereas those on *a*, *b*, . . . are real things and mean something. I would not deny the objection. Although hypothetical unrealities may illuminate the significance of realities, it is the realities that make science. All I was trying to do was to supplement Spearman's discussion of the universality of g with a little contribution on the relativity of g—as might be expected of an erstwhile physicist! It seems to be an undeniable statistical fact that batteries of intelligence tests as given and as scored tend to be what has been termed hierarchical in that they tend to satisfy the tetrad relations (7). This fact means something, it needs to be explained, Spearman has offered an explanation. Possibly the explanation should have laid more emphasis on the tests and less on the general intelligence —I do not know—but in Spearman's system we have a method of examining our data, of discussing its implications, of organizing it into a philosophical system, just as we have in Einstein's, and at least for the immediate future the system propounded in "The Abilities of Man" can not be ignored by those working in its field. That is why I said that the Grote professor of philosophy of mind at the University of London has written an important book. Moreover, it is clearly, spiritedly, suggestively, in places even provocatively, written; intelligible and entertaining even to the general reader. The mathematics has been put out of the way in a highly compressed appendix. I have

chosen to take the risk of misrepresenting the character of the book by writing a very lop-sided review with its emphasis chiefly on the appendix because I know that this has offered difficulties to some very intelligent readers, because it appears logically fundamental to the whole system, and because some of its important logical implications seem not to have been expressed by the author in the main text.

Edwin B. Wilson
Harvard School of Public Health

CONTENTS

Multidimensional Data Representations: When & Why

Ingwer Borg
Editor

MATHESIS PRESS Ann Arbor, Michigan 1981

MATHESIS PRESS
2664 Lowell Road, Ann Arbor, Michigan, 48103

PRINTED IN THE UNITED STATES OF AMERICA