The *mathesis* of intelligence

Hans Varghese Mathews

1

A recent book detailing the 'facts and fallacies' of measuring intelligence describes the basis of its approach thus: "*the core notion of what the word intelligence means is embedded in the language we use*" its author maintains, and "*the purpose of a theory of measurement is to make explicit, and to refine, what is already implicit in the common usage of the word*".[1] The 'theories of measurement' adverted to here are justifications, apparently, for the numerical measures of intelligence that psychometricians might devise, using as data results from tests of intelligence; and such justifications will presumably be offered in the course of "*a dialogue between a proposed measure on the one hand, and the everyday use of the term on the other*". How such a 'dialogue' might proceed is actually sketched out: and the expectation is that 'intuitions' regarding intelligence will countermand the process of testing, when necessary, by identifying those tasks on the tests which are likely to have produced whatever mismatch there may be between intuition and the deliverances of testing.[2] The intuitions of the psychometrician will enjoy the sanction of common usage apparently, and the suggestion is that, through the repeated accomodation of testing to the promptings of common understanding, "*the measurement process itself is helping us define exactly what we mean when we use the word intelligence*". In proceeding so the intent is not "*to eliminate subjectivity*" but, rather, "*to bring the performance of our measuring instrument into line with the commonly accepted meaning*" of the word "intelligence". The 'measuring instruments' here are putative tests of intelligence together with numerical recipes for producing, from appropriate numerizations of an individual's performance on the various tasks set there, a summary number measuring his or her intelligence; and the successful rectification of these instruments through such dialogue will, it is claimed, at the same time "*force us to clarify and, so far as possible, agree on what we mean by intelligence*".

One might wonder if common usage should be so attentive to psychometric practice, and so amenable, as to be refined or even clarified by the sort of exercises just proposed; but the 'dialogue' mooted above should securely identify a good many tasks where intelligence can be exhibited, at any rate, and to that extent might do duty for an 'operational definition' of the term. With operational

[1] MEASURING INTELLIGENCE: FACTS AND FALLACIES by David Bartholomew, Cambridge University Press, 2004. Quoted text of any appreciable length is italicized, so that its extent is easily gauged; and quoted text that is itself in italics will be underlined if the emphasis in the text would otherwise be lost.

[2] See the section DEFINITION BY DIALOGUE in the chapter which ponders 'the end of IQ'. Bartholomew is a considerable figure in the world of psychometry, and a master of its protocols presumably; he admits, though, that such 'dialogues' have seldom taken place in the putting together of intelligence tests.

definitions of its terms in hand the daily business of a science can get underway: experiments can be designed, the apparatus for them assembled, and the results of experiment recorded in unambiguous ways. The controversy attending the measurement of intelligence suggests that common usage, in English at least, will resist whatever operational definitions of "intelligence" psychometricians might propose.[3] But the dialogue our text proposes should allow the business of measuring intelligence to proceed nonetheless: whether or not common usage is clarified or refined thereby. Now the primary condition the measurement of intelligence must meet to be 'scientific' at all, or so psychometric practice suggests, is that any numerical measure of intelligence be derived in some methodical way from *nothing more* than individuals' numerized performances on the various tasks set by some putative test of intelligence: no other sorts of 'raw' given or 'cooked' constraint should enter into its determination. Any measure of intelligence must be determined by an *empirical model* which has been generated from the performances of a representative sample of human beings on some suitable test of intelligence: and so will take as given only their performances, numerized as scores, on the various tasks set by that test of intelligence. Such models are the numerical recipes of the 'measuring instruments' above, of course; they typically yield their summary measures by somehow weighting and summing some tractable transformations of these scores, and the epithet "empirical" insists that no prior constraint be placed on how these weights relate to one another. Empirical models are supposed to 'let the data speak for themselves' as much as they can; and the extent to which measures of intelligence can be kept from going beyond their data should come clear when the sort of model most in use is examined.

Empirical models of intelligence are nowadays constructed using a statistical technique that psychometricians had pioneered, actually, to get more informative measures than the usual 'intelligence quotients' will provide. The technique yields measures of 'general intelligence' which may be decomposed as weighted sums of measures of specific mental abilities: which are themselves supposed to be tested by distinct ensembles within the collection of tasks set by a test of intelligence. These distinguishing ensembles of tasks would be *discovered* in the process of constructing the model, ideally, and not specified beforehand: the data are supposed to disclose these specific abilities, and that is a further 'empirical' feature of such models. Specific mental abilities and general intelligence are both supposed to be manifested in various ways by individuals' performances on the tasks set by a test of intelligence, then, and the scores that individuals achieve on these tasks are taken for observations of a set of interrelated *manifest* variables, within which indications of the specific abilities and general intelligence they would varyingly possess are *latent*. The technique used to at once discern and measure these latent features is called factor analysis;

[3] "*One would have thought*" our author says, seeming ingenuous momentarily, "*that the measurement of intelligence would have been one of those worthy but unexciting statistical activities — like the production of the Retail Price Index — vital for the well-being of society but best left to those expert in these things. But this is not so, and the reason*", he quickly remembers, "*is that the whole question is bound up with our nature as persons*".

the technique "*attempts to mimic the mental processes by which we arrive at the concept of an underlying (latent) variable (or variables) influencing a set of manifest variables*" allegedly, and that is what seems to recommend it.

Putting aside for the moment the question of how concepts are arrived at, it seems true enough that human beings routinely assess what may be called latent features of objects, or events or situations, by differentially taking account of their manifest features. Gauging someone's attitude toward something from what they do or say is an obvious example now of assessing a latent thing by somehow weighting manifest ones; but "*even in such everyday matters as choosing from a menu*" we are supposed to "*balance our likes and dislikes*" in a way that factor analysis 'attempts to mimic'. One wonders how like such a balancing the social process of concepts being arrived at could be; but we cannot let that question detain us. Assessing the 'overall size' of a familiar object, like a house or a tree, is an example the text dwells on at some length. We are often called upon, and may often find it useful, to describe houses or trees as large or middling or small. In making mutually acceptable assessments of overall size — which would be a 'latent' quantity compared to the 'manifest' number and extent of visible features like doors and windows, or trunks and branches, and such like — different individuals presumably group and weight the manifest features of houses and trees in characteristic and congruent ways: and the claim here seems to be that the process which determined the prevalent grouping and differential weighting is mirrored, somehow or other, by how variables are grouped and weighted through factor analysis.

This seems to be a claim about how words that name latent features gain their meanings. If we take seriously Wittgenstein's dictum that 'meaning is use', then the meaning of a word that names a latent feature would now *consist in* some particular and socially prevalent grouping and differential weighting of the manifest features of those objects or events or situations that are thought to possess the latent feature the word names; and, to state it again, the surprising claim now is that the social process which determines the prevalent grouping and weighting of such features is mirrored by factor analysis. The suggestion that factor analysis 'mimics the mental processes by which we arrive at the concept of a latent variable' squares nicely with all this, of course, and perhaps one was meant to gloss it so. Anyhow, if the deriving of measures formally follows what naturally happens as the words gain their primary or 'core' meaning, then the process of measuring intelligence would constitute a mathesis — a mode of disciplined learning — that indeed refines and clarifies common understandings of the word "intelligence". The theory of measurement advertised above would comprehend this mathesis, and disclose the rationale of its protocols; and that factor analysis formally mirrors the natural formation of meaning seems a postulate principal to theory here.

We shall need some account of factor analysis to proceed; and one that is technical enough to adumbrate and assess the large claim just set out. The next section is a primer on factor analysis; and its readers are assumed to possess the rudiments of multivariate statistics. There are informal descriptions of factor analysis posted on the Web, which those to whom such matters are new may consult; but such readers should be able to decipher what follows if they acquaint themselves a little, as they proceed, with the technical terms they meet.[4] The exercise may not be pleasant for those who are unfamiliar with mathematical notation; but I hope that such readers are not deterred on that account, and that they will go some way with me through the next section; and I ask them to keep in mind that the notation only abbreviates sentences in a language whose grammar, compared to that of most natural languages, is elementary.[5]

2

Let $X_1, X_2, \ldots, X_p$ be jointly distributed and real-valued random variables which, in that order, comprise the p components of a random vector $\mathbf{X}$. We assume for simplicity that the mean or 'expected value' of each X_j is 0, and abbreviate that with $E(\mathbf{X}) = \mathbf{0}$ as usual.[6] Suppose that, for some $m < p$, the components of $\mathbf{X}$ can be reconstituted, with some residue, from m many variable quantities $F_1, F_2, \ldots, F_m$ comprising in that order a random vector $\mathbf{F}$, in the following sense: we suppose that $\mathbf{X} = Q\mathbf{F} + \boldsymbol{\varepsilon}$ where Q is a $p \times m$ coefficient matrix of rank m, and $\boldsymbol{\varepsilon}$ a random vector satisfying $E(\boldsymbol{\varepsilon}) = \mathbf{0}$ and having ordered components $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_p$ which are independent of each other and of all the $\{F_k\}$ as well. We must have $E(\mathbf{F}) = \mathbf{0}$ now; the last two constraints may be expressed as $E(\varepsilon_i \varepsilon_j) = 0$ when $i \neq j$ and $E(\mathbf{F} \cdot \boldsymbol{\varepsilon}^{\tau}) = \mathbf{0}$.

---

[4] There are a number of sites on the Web which introduce readers to mathematical terms: the site maintained by Wolfram Research is one such, the Wikipaedia another. But some acquaintance with the elementary notions of linear algebra is all that is really needed here; readers to whom matrices and vectors and the usual operations with them are not unfamiliar should be able to gloss what follows well enough, after getting themselves around some elementary statistical ideas.

[5] We note a particular notational liberty taken to keep the text from looking too cluttered. A collection $O_1, O_2, \ldots, O_n$ of objects is collectively referred to with $\{O_i\}_{i=1}^{n}$ usually; we write $\{O_i\}$ simply, because the values which the index i takes can always be established from the context here. Different symbols will be used for indices when different collections are being referred to thus; and the context will always make clear the distinct values these indices can take. I use the old-fashioned M^{τ} for the transpose of a matrix M. A vector $\mathbf{v}$ is for notational convenience regarded as a matrix having one column and as many rows as it has components; its transpose $\mathbf{v}^{\tau}$ is then a matrix with one row and as many columns as components. A matrix with m columns will sometimes be specified here as a concatentation $[\mathbf{v}_1 \,|\, \mathbf{v}_2 \,|\, \ldots \,|\, \mathbf{v}_m]$ of the m vectors which its columns are; when M_1 and M_2 are matrices with the same number of rows we write $[M_1 \,|\, M_2]$ for the matrix obtained by adjoining, on the right, the columns of M_2 to those of M_1. With particular reference to what follows note that the matrix product $\mathbf{v} \cdot \mathbf{w}^{\tau}$ of a p-vector $\mathbf{v}$ and a q-vector $\mathbf{w}$ is a $p \times q$ matrix, which is sometimes called the 'outer' product; the usual 'inner' or 'dot' product of p-vectors $\mathbf{u}$ and $\mathbf{v}$ is now the matrix product $\mathbf{u}^{\tau} \cdot \mathbf{v}$. An *eigenvector* $\mathbf{e}$ of a matrix M has unit length and satisfies $M \cdot \mathbf{e} = \lambda \mathbf{e}$ for some scalar *eigenvalue* λ; a *positive definite* matrix has only positive such λ. The *trace* $tr(M)$ of a square matrix M is the sum of its diagonal elements. A vector or matrix identity asserts the identity of corresponding components. Variable vectors or matrices have scalar variables for components; the expected value of a vector or matrix variable is the vector or matrix of the expected values of their components.

[6] Otherwise we simply replace $\mathbf{X}$ with $\mathbf{X} - E(\mathbf{X})$. Note that $\mathbf{0}$ here always denotes the vector or matrix whose each component is 0.

The $\{X_j\}$ here are observed or *manifest* variables, while the $\{F_k\}$ are *latent factors* imputed to the process manifested through the former; each ε_j is correlated with X_j alone, and on that account is called a *specific factor,* while the components of $\mathbf{F}$ are called *common factors*: each of which could be correlated, possibly, with any of the manifest variables. Let $\boldsymbol{\Sigma}$ be the covariance matrix $E(\mathbf{X}\cdot\mathbf{X}^\tau)$ of $\mathbf{X}$ and $\boldsymbol{\Phi}$ the covariance matrix $E(\mathbf{F}\cdot\mathbf{F}^\tau)$ of $\mathbf{F}$; let ψ_j be the variance of ε_j and write $\boldsymbol{\Psi}$ for the diagonal matrix $E(\boldsymbol{\varepsilon}\cdot\boldsymbol{\varepsilon}^\tau)$ which assembles these specific variances along its diagonal. The relation $\mathbf{X} = Q\mathbf{F} + \boldsymbol{\varepsilon}$ implies that $\boldsymbol{\Sigma} = Q\Phi Q^\tau + \Psi$.[7] Being the covariance of $\mathbf{X} - \boldsymbol{\varepsilon}$ the matrix $\boldsymbol{\Sigma} - \boldsymbol{\Psi} = Q\Phi Q^\tau$ records the correlations between the $\{X_j\}$ that are *due to* $\mathbf{F}$ one might say; and estimating it is a basic step in factor analysis.

Now we have $X_j = \sum_{l=1}^{m}(Q)_{jl}F_l + \varepsilon_j$ for any given j and so, as ε_j and any other F_k are uncorrelated, the covariance of any pair X_j and F_k of manifest variable and latent factor is the jk-th entry $\sum_{l=1}^{m}(Q)_{jl}\cdot(\Phi)_{lk}$ of the matrix $Q\Phi$. The factors can be told from each other only by how differently they correlate with the variables, but these covariances will not by themselves let us compare how differently a variable X_j is correlated to different factors; one needs the variances of the $\{X_j\}$ and the $\{F_k\}$ for that. The former can be estimated from the data; the latter must be reckoned otherwise, and to avoid that complication the convention is to endow all latent factors with unit variance. A factor is said to 'load' a variable to the extent of their correlation; corrected with the variances of the $\{X_j\}$ the entries of $Q\Phi$ constitute the matrix of *factor loadings,* each column of which identifies a latent factor through its correlations with the manifest variables.

To carry out a factor analysis of $\mathbf{X}$ one must have a sufficient number of observations of its components, gathered so as to estimate $\boldsymbol{\Sigma}$ reliably. The factor analyst must decide upon the number m of imputed common factors, and then estimate the coefficients Q and the covariance Φ. The estimation of Ψ is a preliminary to either the first or the second of these steps, depending on the method preferred: we shall come to that shortly. In practice, as we shall see, estimating the matrix of factor loadings is the crucial business; this is done by estimating the product $Q\Phi$ first, but in a manner which allows one to estimate Φ itself, and hence Q as well. The estimate of the product $Q\Phi$, corrected by dividing each jk-th entry with the estimated variance of X_j in order to estimate the correlation between each pair X_j and F_k, yields the estimated loadings; and the putative latent factors are then describable in terms of the manifest variables.[8]

Estimating the loadings is simpler if we assume that Φ is the product $\Theta\Theta^\tau$ of some invertible $m \times m$ matrix Θ and its transpose; which obtains if Φ is positive definite, for instance. That is not a restrictive assumption: the

[7] Because $E(\mathbf{X}\cdot\mathbf{X}^\tau) = E((Q\mathbf{F}+\boldsymbol{\varepsilon})\cdot(Q\mathbf{F}+\boldsymbol{\varepsilon})^\tau) = E((Q\mathbf{F}+\boldsymbol{\varepsilon})\cdot(\mathbf{F}^\tau Q^\tau + \boldsymbol{\varepsilon}^\tau))$, which reduces to $Q\cdot E(\mathbf{F}\cdot\mathbf{F}^\tau)\cdot Q^\tau + E(\boldsymbol{\varepsilon}\cdot\mathbf{F}^\tau)\cdot Q^\tau + Q\cdot E(\mathbf{F}\cdot\boldsymbol{\varepsilon}^\tau) + E(\boldsymbol{\varepsilon}\cdot\boldsymbol{\varepsilon}^\tau) = Q\Phi Q^\tau + \Psi$ since $E(\mathbf{F}\cdot\boldsymbol{\varepsilon}^\tau) = E(\boldsymbol{\varepsilon}\cdot\mathbf{F}^\tau) = \mathbf{0}$.

[8] When the units of the manifest variables are incommensurable the $\{X_j\}$ are replaced by their standardizations before a factor analysis is performed; in that case the estimate of $Q\Phi$ is already a matrix of estimated correlations.

covariance of a random vector is positive definite provided no component is a linear combination of the others. We have $\mathbf{\Sigma} - \Psi = Q\Phi Q^{\tau} = Q\Theta\Theta^{\tau}Q^{\tau}$ then. Being of maximal rank the columns of Q serve as a basis for some subspace V of dimension m in $\mathbf{R}^p$; then the columns of $\Lambda = Q\Theta$ serve as a basis for V as well. We have $\Lambda\Lambda^{\tau} = Q\Theta\Theta^{\tau}Q^{\tau} = \mathbf{\Sigma} - \Psi$ now, and $Q\Phi = \Lambda\Theta^{-1}\Theta\Theta^{\tau} = \Lambda\Theta^{\tau}$ yields the matrix of loadings for $\mathbf{F}$. The loadings are obtained by *rotating* Λ with Θ one says.[9] Note that the columns of $Q\Phi = \Lambda\Theta^{\tau}$ also provide a basis for V : which is often called the *factor space.*

The usual factor analytic practice is to obtain a matrix L of maximal rank where the product LL^{τ} approximates some reliable estimate of $\mathbf{\Sigma} - \Psi$ sufficiently, and to assume that the columns of L do indeed provide a basis for the factor space V that Q determines. The rank of L estimates the number m of latent factors now; the sample covariance matrix of the observed data is used to estimate $\mathbf{\Sigma}$ and, as we just noted, Ψ is either estimated beforehand or estimated along with L. We shall take up shortly the matter of deciding when LL^{τ} estimates $\mathbf{\Sigma} - \Psi$ sufficiently. A suitable L can be rotated to obtain an estimate of $Q\Phi$, and hence of the loadings for $\mathbf{F}$, when the columns of L are taken to provide a basis for V, since the columns of $Q\Phi$ make such a basis; the problem now is to choose an appropriate T to rotate L with, and this brings us to what is most controversial about factor analysis.

The data themselves cannot direct the choice of a rotating T at all, since they suggest nothing about Q and Φ. In practice T is chosen so that the rotated matrix LT^{τ} is easily interpretable as a matrix of loadings, after correction with the estimated variances of the $\{X_j\}$, by the factor analyst who chooses: interpretation being easy to the extent that factors can be readily described by how differently they correlate with the manifest variables. But loadings that one factor analyst finds easy to interpret may seem opaque to another, and conversely, so 'ease of interpretation' is not a criterion that can be made formal. Having chosen T one takes $\widehat{\Phi} = TT^{\tau}$ as one's estimate of the covariance Φ, and $\widehat{Q} = LT^{-1} = LT^{\tau}(TT^{\tau})^{-1}$ as one's estimate of the coefficient matrix Q then, since one obtains $\widehat{Q}\widehat{\Phi} = LT^{\tau}$ with only that choice now. Because $\widehat{Q}\widehat{\Phi}\widehat{Q}^{\tau} = LT^{-1} \cdot TT^{\tau} \cdot (T^{\tau})^{-1}L^{\tau} = LL^{\tau}$ we have $\widehat{Q}\widehat{\Phi}\widehat{Q}^{\tau}$ estimating $\mathbf{\Sigma} - \Psi$ sufficiently for *any* choice of T when LL^{τ} itself does so; and since the only constraint that the data place on the estimates $\widehat{Q}$ and $\widehat{\Phi}$ is that $\widehat{Q}\widehat{\Phi}\widehat{Q}^{\tau}$ estimate $\mathbf{\Sigma} - \Psi$ sufficiently, we see that the data themselves cannot decide between rotations to estimate loadings. The factor analyst goes beyond the data, therefore, when choosing matrices of loadings that specify latent factors in terms of manifest variables: and this makes controversial the use of factor analysis in the construction of empirical models.

[9] Here is the rationale for the phrasing. Suppose $\Lambda = [\mathbf{v}_1|\mathbf{v}_2|\ldots|\mathbf{v}_m]$ and $\Theta = (r_{kl})$; let $f : \mathbf{R}^m \longrightarrow \mathbf{R}^p$ be the linear map taking each standard basis vector $\mathbf{e}_l$ to $\mathbf{v}_l$; now $\Theta{\cdot}\mathbf{e}_l = \sum_k r_{kl}\mathbf{e}_k$, so rotating the standard basis of $\mathbf{R}^m$ with Θ and then applying f gives us the mapping $\mathbf{e}_l \longrightarrow \sum_k r_{kl}\mathbf{v}_k$. But $\Lambda\Theta^{\tau}$ is $[\sum_k r_{k1}\mathbf{v}_k \mid \sum_k r_{k2}\mathbf{v}_k \mid \ldots \mid \sum_k r_{km}\mathbf{v}_k]$ precisely: so the columns of $\Lambda\Theta^{\tau}$ are the vectors we would get if we were to map the columns of Λ back to corresponding standard basis vectors in $\mathbf{R}^m$, rotate these with Θ, and then map the rotated basis back to $\mathbf{R}^p$ using f.

The entries of the loading matrix measure the extent to which a variable manifests or expresses a factor, one could say, and the variation between the estimated loadings of variables by a putative factor is what will allow us to describe that factor in terms of the manifest variables. A clutch of manifest variables that each express some one putative factor greatly while expressing all the other factors negligibly provides a ready description of that factor; such selectively loaded variables are sometimes called *perfect indicators,* and clutches of perfect indicators are called *perfect clusters.* Were all the latent factors selectively expressed so, by clusters of perfect indicators, the matrix of loadings could be readily interpreted. The circumstance that each indicator in a perfect cluster would be highly correlated with every other, and correlated with any other manifest variable only to the extent that the factors they separately express were correlated, would readily distinguish the factors.[10] We cannot generally suppose, however, that perfect indicators exist; each manifest variable might express more than one factor substantially, and some might not express any one factor substantially more than any other. Nonetheless, the desideratum that manifest variables expressing some one latent factor more than any other be discerned, when such exist, seems to usually direct the process of rotation.[11] But there seems to be no way of assessing, from a random sample of observations, the likelihood that there are perfect indicators in a given collection of jointly distributed variables; or of assessing the chance that any variable may be such. So the specification of latent factors, as that is usually done, can only be tentative: unless the manifest variables happen to form highly intercorrelated and mutually uncorrelated groups, in which case the factors identified by each group will have little to do with each other.

We turn next to the matter of making an initial choice L whose columns may be taken, when LL^τ estimates $\mathbf{\Sigma} - \mathbf{\Psi}$ sufficiently, as a basis for the factor space that the coefficient matrix Q determines. There appear to be three methods of obtaining L in common use; we shall very summarily describe them. The method of *principal factors* forms an estimate $\widehat{\mathbf{\Psi}}$ of the specific variances $\mathbf{\Psi}$ first, usually estimating each ψ_l by regressing X_l against the remaining $\{X_j\}$, and, using the sample covariance matrix $\widehat{\mathbf{\Sigma}}$ to estimate $\mathbf{\Sigma}$, takes the substantially positive eigenvectors of $\widehat{\mathbf{\Sigma}} - \widehat{\mathbf{\Psi}}$ as the columns of

[10] The correlation between any two factors would be appreciably less, usually, than correlations within the clusters of perfect indicators that separately express them. The matrix of loadings would have a particular structure now: each of its rows would contain one number appreciably larger than all the other in that row, and each column would contain a collection of entries that were comparable between themselves, in size, and appreciably larger than the other entries.

[11] The standard varimax method tries to maximize the variance in each column of the loading matrix, for example, while the direct quartimin method tries to maximize the variance in each of its rows; both strategies of rotation seem designed to detect such manifest variables. Other strategies of rotation are described in *An Overview of Analytic Rotation* by Micheal Browne, published in *Multivariate Behavioural Research 36(1)* in 2001. This review leaves one with the distinct impression that the digital automation of factor analysis, which enables users of statistical software to perform them readily, has actually made rotating towards loadings a less flexible practice, generally, than it used to be.

L, after scaling each such with the square root of its associated eigenvalue.[12] The method of *principal components* does as much with the eigenvectors of $\widehat{\mathbf{\Sigma}}$ itself, and uses the diagonal of $\widehat{\mathbf{\Sigma}} - LL^\tau$ as its estimate of $\mathbf{\Psi}$. The method is called so because the eigenvectors $\{\mathbf{u}_j\}$ of a positive definite $\mathbf{\Sigma}$ yield a collection $\{Y_j = \mathbf{u}_j^\tau \cdot \mathbf{X}\}$ of independent random variables with the variance of each Y_j being the eigenvalue associated to $\mathbf{u}_j$, and these variables are called the principal components of $\mathbf{X}$. The rationale for these proceedings will come clear in a moment. [13] These ways of obtaining L make no assumptions about how the $\{X_j\}$ are jointly distributed. When that distribution may be supposed normal L can be obtained as a 'maximum likelihood' estimate, together with a like estimate $\widehat{\mathbf{\Psi}}$ of $\mathbf{\Psi}$, with $LL^\tau + \widehat{\mathbf{\Psi}}$ being a maximum likelihood estimate of $\mathbf{\Sigma}$ now. In this last case, and when the size of the sample is large, in itself and compared to the number p of manifest variables, there is a 'significance test' for how well $LL^\tau + \widehat{\mathbf{\Psi}}$ might estimate $\mathbf{\Sigma}$. With L and $\widehat{\mathbf{\Psi}}$ that pass this test the trace of LL^τ should be a substantial fraction of the trace of $\widehat{\mathbf{\Sigma}} - \widehat{\mathbf{\Psi}}$, and LL^τ should well approximate $\widehat{\mathbf{\Sigma}}$ off their respective diagonals. With the method of principal factors the suitability of L is tested by precisely these considerations.[14] With the method of principal components it is usual to insist that the trace of LL^τ be a substantial fraction of the trace of $\widehat{\mathbf{\Sigma}}$ itself and, again, that LL^τ well approximate $\widehat{\mathbf{\Sigma}}$ off their respective diagonals.

The trace of $\mathbf{\Sigma}$ measures the *total variance* of $\mathbf{X}$, computed as the expected value of the squared length $||\mathbf{X}||^2$ here, while the trace of $\mathbf{\Sigma} - \mathbf{\Psi}$ measures the total variance of $\mathbf{X}-\boldsymbol{\varepsilon}$. The eigenvalues of $\mathbf{\Sigma}$ and $\mathbf{\Sigma} - \mathbf{\Psi}$ sum to their respective traces; and the trace of LL^τ sums the eigenvalues corresponding to the eigenvectors chosen and scaled to construct L.[15] The trace of $\mathbf{\Sigma} - \mathbf{\Psi}$ may be thought to measure the total variance in $\mathbf{X}$ that is due to the common factors, and is called the *common variance* therefore; as entries off the diagonals of both these matrices measure the covariance of $\mathbf{X}$, we could summarize the demand on L by saying that LL^τ must sufficiently conserve both the covariance and the common variance of $\mathbf{X}$. Though the adequate conserving of common variance determines how L is chosen when either principal factors

---

[12] A 'positive eigenvector' is simply one associated to a postive eigenvalue. The specific variance ψ_l is usually estimated with $1 - R_l^2$ where R_l^2 is the squared multiple correlation coefficient obtained by regressing X_l against the remaining $\{X_j\}$.

[13] The salient circumstance here is that a postive definite matrix A can be decomposed as a product $RDR^\tau = (RD^{1/2})(RD^{1/2})^\tau$ where R is a matrix whose columns are eigenvectors of A and D a diagonal matrix whose non-zero entries are its eigenvalues, with $(D)_{jj}$ being the eigenvalue associated to the eigenvector that constitutes the j-th column of R.

[14] Suppose the substantially positive eigenvectors of $\widehat{\mathbf{\Sigma}} - \widehat{\mathbf{\Psi}}$ do not yield a suitable L with the choice of $\widehat{\mathbf{\Psi}}$ initially made. Let L_0 denote the matrix whose columns consist of these eigenvectors; the estimation of $\mathbf{\Sigma} - \mathbf{\Psi}$ could be repeated using the diagonal elements of $L_0 L_0^\tau$ to re-estimate $\mathbf{\Psi}$ now, with $(\widehat{\mathbf{\Sigma}})_{ii} - (L_0 L_0^\tau)_{ii}$ being the new estimate of ψ_i in fact. Let $\mathbf{\Psi}_1$ denote this re-estimate of $\mathbf{\Psi}$, and let L_1 be the matrix consisting of the substantially positive eigenvectors of $\widehat{\mathbf{\Sigma}} - \mathbf{\Psi}_1$ now; the suitability of L_1 with regard to $\widehat{\mathbf{\Sigma}}$ and $\mathbf{\Psi}_1$ may be assessed as above. The method of *iterated principal factors* continues this process of restimation until the suitability of the repeated estimates L_* with regard to $\widehat{\mathbf{\Sigma}}$ and the repeated estimates $\mathbf{\Psi}_*$ cannot be improved.

[15] If $A = RDR^\tau$ with $RR^\tau = I$, as with $\mathbf{\Sigma}$ and $\mathbf{\Sigma} - \mathbf{\Psi}$ here, then $tr(A) = tr(D)$.

or principal components are employed, the adequate conserving of covariance will usually follow thereon as well.[16] The norms which secure agreement on what would count as 'substantial' and 'well approximating' here seem to vary somewhat between regions of social science.[17]

Suppose the estimated number m_1 of common factors remains substantial compared to the number p of manifest variables, after a first analysis of a sample $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ of $\mathbf{X}$, with the correlations between many pairs of imputed factors seeming significant; one could repeat what one did with the imputed factors $\mathbf{F}$ now, using the estimated covariance of $\widehat{\Phi}$ of $\mathbf{F}$ just as one had used the estimate $\widehat{\mathbf{\Sigma}}$ before, and using the *factor scores* $\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_n$ as one had used the observations $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ before, where $\mathbf{f}_i$ estimates the expected value of $\mathbf{F}$ given that $\mathbf{X} = \mathbf{x}_i$.[18] Correlations between factors would measure their 'imperfection' when rotation was done with a view to finding perfect clusters, one would think, and repeating the process seems warranted then. A factor analysis performed on the factors first obtained, which are called factors of order 1 now, would yield factors of order 2 which load the factors of order 1 in particular ways. One can now estimate how these latter factors might factorize the mainfest variables themselves, and so estimate their loadings by these factors.[19]

[16] Let S denote $\widehat{\mathbf{\Sigma}}$ or $\widehat{\mathbf{\Sigma}} - \widehat{\Psi}$; let $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_p$ and $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p$ be the eigenvectors and corresponding eigenvalues of S. When $\mathbf{X}$ has been randomly sampled S should have as many distinct eigenvectors as either $\mathbf{\Sigma}$ or $\mathbf{\Sigma} - \Psi$. Now $L = [\sqrt{\lambda_1}\mathbf{e}_1 \,|\, \sqrt{\lambda_2}\mathbf{e}_2 \,|\, \ldots \,|\, \sqrt{\lambda_m}\mathbf{e}_m]$ for some $m < p$ with $\lambda_1 + \lambda_2 + \ldots + \lambda_m$ being a substantial part of $\lambda_1 + \lambda_2 + \ldots + \lambda_p$. With $C = [\mathbf{e}_1 | \mathbf{e}_2 | \ldots | \mathbf{e}_p]$ and Λ the $p \times p$ diagonal matrix where $(\Lambda)_{jj} = \lambda_j$ we have $S = C \Lambda C^{\tau}$; so $(S)_{jk} = \sum_{r=1}^{p} \lambda_r e_{jr} e_{kr}$ where e_{lr} is the rth component of $\mathbf{e}_l$, while $(LL^{\tau})_{jk} = \sum_{r=1}^{m} \lambda_r e_{jr} e_{kr}$. Since $|e_{lr}| \leq 1$ always, the difference between these sums will depend on the quantities $\lambda_{m+1}, \lambda_{m+2}, \ldots \lambda_p$; when these are all appreciably less than 1, for instance, that difference should be small.

Note that when L is determined by the eigenvectors and eigenvalues of $\widehat{\mathbf{\Sigma}}$ the trace of LL^{τ} estimates the total variance of the random vector obtained by projecting $\mathbf{X}$ on to the subspace determined by L's columns; where the eigenvectors and eigenvalues of $\widehat{\mathbf{\Sigma}} - \widehat{\Psi}$ determine L the trace of LL^{τ} would estimate the total variance of the random vector one would obtain by projecting $\mathbf{X} - \boldsymbol{\varepsilon}$ on to the subspace determined by L. Geometrically considered, then, factor analysis consists in specifying a hyperplane which conserves the common variance of $\mathbf{X}$ when projected upon, and in choosing an appropriate set of cardinal directions for this hyperplane.

[17] Much seems to depend on what amount of correlation between variables would *generally* count as significant; from Ian Deary's INTELLIGENCE: A VERY SHORT INTRODUCTION one gathers that in psychometry the correlation coefficient need only come to $1/3$ or so for two variables to be regarded as sufficiently correlated.

[18] Statistical opinion seems divided on what the best way of estimating expected factor values may be; but that is not thought a grave matter. One common method is to solve the equation $\mathbf{x}_i = \widehat{Q}\mathbf{f}_i + \widehat{\Psi}$ for each $i \in \{1, 2, \ldots n\}$ using *weighted least squares* regression. Any method of obtaining factor scores must ensure, of course, that the sample covariances computed with them agree closely with the entries of the estimated covariance $\widehat{\Phi}$ of $\mathbf{F}$.

[19] Suppose $\mathbf{X} = Q_1\mathbf{F}_1 + \boldsymbol{\varepsilon}_0$ and $\mathbf{F}_1 = Q_2\mathbf{F}_2 + \boldsymbol{\varepsilon}_1$ are factorizations of $\mathbf{X}$ and $\mathbf{F}_1$, by $\mathbf{F}_1$ and $\mathbf{F}_2$ respectively. We have $\mathbf{X} = Q_1 Q_2 \mathbf{F}_2 + Q_1\boldsymbol{\varepsilon}_1 + \boldsymbol{\varepsilon}_0$ now; but since the components of $Q_1\boldsymbol{\varepsilon}_1 + \boldsymbol{\varepsilon}_0$ may be intercorrelated, we may not have a factorization of $\mathbf{X}$ by $\mathbf{F}_2$ here. Assuming that the correlation matrix of $Q_1\boldsymbol{\varepsilon}_1 + \boldsymbol{\varepsilon}_0$ has positive eigenvalues, let $\tilde{\boldsymbol{\varepsilon}} = P^{\tau} \cdot (Q_1\boldsymbol{\varepsilon}_1 + \boldsymbol{\varepsilon}_0)$ be the vector of its principal components, with P being the orthogonal matrix of eigenvectors. Now $P^{\tau}\mathbf{X} = P^{\tau} Q_1 Q_2 \mathbf{F}_2 + \tilde{\boldsymbol{\varepsilon}}$ gives us a factorization of

What is called *hierarchical factor analysis* continues with such repetitions until the imputed latent factors have been reduced to an explanatorily manageable number, or until they appear to be uncorrelated; and the factor analysis of intelligence proceeds so, in fact. The performance of a randomly chosen collection of individuals on some putative test of intelligence provide the initial observations, usually after some statistical preparation. The test will consist of a number of tasks, each of which will typically have to be essayed a number of times in taking the test. Since the results of a factor analysis are secure to the extent that the covariance of the manifest variables determines their joint distribution, and since the most tractable such distribution is the normal one, aggregating the performances on each task so as to obtain somewhat normally distributed variables is a desideratum acknowledged by factor analysts of intelligence. Some form of the Central Limit Theorem would be resorted to now, and the manner of aggregation would not compromise too much, or so one imagines, the 'manifest' character of the resulting variables: which now numerize as scores individual performances on a number of distinct tasks.

Factor analysts of intelligence sometimes collate and transform in a particular way, due to Schmid and Leiman, the successive ensembles of factors identified by a hierarchical factor analysis. Suppose $\mathbf{F}_1, \mathbf{F}_2, \ldots, \mathbf{F}_r$ are successive such factorizations of our random vector $\mathbf{X}$, where the components of

---

$\mathbf{Y} = P^\tau \mathbf{X}$ by $\mathbf{F}_2$. Since $E(\mathbf{Y}\cdot\mathbf{F}_2^\tau) = P^\tau \cdot E(\mathbf{X}\cdot\mathbf{F}_2^\tau)$ relates the covariance of $\mathbf{Y}$ and $\mathbf{F}_2$ to the covariance of $\mathbf{X}$ and $\mathbf{F}_2$, we can derive the correlations of $\mathbf{X}$ with $\mathbf{F}_2$ from how these factors load $\mathbf{Y}$: with Φ_2 being the covariance of $\mathbf{F}_2$, we get $E(\mathbf{X}\cdot\mathbf{F}_2^\tau) = Q_1 Q_2 \Phi_2$ from $P^\tau \cdot E(\mathbf{X}\cdot\mathbf{F}_2^\tau) = E(\mathbf{Y}\cdot\mathbf{F}_2^\tau) = P^\tau Q_1 Q_2 \Phi_2$.

Writing Ψ_0 for the specific variance $E(\boldsymbol{\varepsilon}_0\cdot\boldsymbol{\varepsilon}_0^\tau)$ now, suppose we had satisfactorily estimated Q_1 as $L_1 T_1^{-1}$ with some rotation T_1 and some L_1 where $L_1 L_1^\tau$ had estimated $\boldsymbol{\Sigma} - \Psi_0$ sufficiently. Now $\widehat{\Phi}_1 = T_1 T_1^\tau$ is the estimate of the covariance $E(\mathbf{F}_1\cdot\mathbf{F}_1^\tau)$ we obtained in modelling the factorization of $\mathbf{X}$ by $\mathbf{F}_1$; let $\widehat{\Psi}_1$ estimate the specific variance of $\mathbf{F}_1$ in the factorization of $\mathbf{F}_1$ by $\mathbf{F}_2$. Suppose next that L_2 is a matrix with $L_2 L_2^\tau$ approximating $\widehat{\Phi}_1 - \widehat{\Psi}_1$ sufficiently; to model the factorization of $\mathbf{F}_1$ by $\mathbf{F}_2$ we would rotate L_2 with some T_2 to get our estimate $L_2 T_2^\tau$ of the covariance $E(\mathbf{F}_1\cdot\mathbf{F}_2^\tau) = Q_2\Phi_2$, taking $T_2 T_2^\tau$ as our estimate of Φ_2. Our estimate of $E(\mathbf{X}\cdot\mathbf{F}_2^\tau) = Q_1 Q_2 \Phi_2$ is $L_1 T_1^{-1} L_2 T_2^\tau$ now: and we could regard ourselves as having rotated the matrix $L_1 T_1^{-1} L_2$ by T_2 in obtaining this estimate. To get correlations of $\mathbf{X}$ by $\mathbf{F}_2$ which are more scrutable as possible loadings we could choose another and more appropriate $\widetilde{T}_2$ to rotate $L_1 T_1^{-1} L_2$ with: and as dividing each $(L_1 T_1^{-1} L_2 \widetilde{T}_2^\tau)_{jk}$ with the estimated variance $\widehat{\sigma}_{jj}$ of X_j estimates the correlations of $\mathbf{X}$ with $\mathbf{F}_2$, we would actually rotate the product $D \cdot L_1 T_1^{-1} L_2$ where D is the diagonal matrix with $(D)_{jj} = 1/\widehat{\sigma}_{jj}$. Supposing that some $\mathbf{X} = Q\mathbf{F}_2 + \boldsymbol{\varepsilon}$ is the actual factorization of $\mathbf{X}$ by $\mathbf{F}_2$, we could take $L_1 T_1^{-1} L_2 \widetilde{T}_2^\tau$ as our estimate of $Q\Phi_2 = E(\mathbf{X}\cdot\mathbf{F}_2^\tau)$, with $\widetilde{T}_2 \widetilde{T}_2^\tau$ being our estimate $\widehat{\Phi}_2$ of Φ_2 now. Then $L_1 T_1^{-1} L_2 \widetilde{T}_2^{-1} = L_1 T_1^{-1} L_2 \widetilde{T}_2^\tau \widehat{\Phi}_2^{-1}$ would be our estimate $\widehat{Q}$ of Q, and $\widehat{\boldsymbol{\Sigma}} - \widehat{Q}\widehat{\Phi}_2\widehat{Q}^\tau$ would be our estimate of the specific variance $\Psi = E(\boldsymbol{\varepsilon}\cdot\boldsymbol{\varepsilon}^\tau)$; and with these we could, of course, estimate the expected value of $\mathbf{F}_2$ given that $\mathbf{X} = \mathbf{x}_i$, for each observation index $i \in \{1, 2, \ldots, n\}$.

The assumption about $Q_1\boldsymbol{\varepsilon}_1 + \boldsymbol{\varepsilon}_0$ on which this factorization of $\mathbf{X}$ by $\mathbf{F}_2$ depends is not easy to examine; but it seems benign. From the stray references I have, the factoring of manifest variables by factors of factors seems a nonce affair, considered generally. But in the particular situtation we are considering, as we shall shortly see, factors of factors have come to be treated in a definite way.

the final ensemble $\mathbf{F}_r$ are independent factors. Let $\widehat{\Phi}_1, \widehat{\Phi}_2, \ldots, \widehat{\Phi}_r \approx \mathbf{I}$ be the successively estimated correlation matrices of these ensembles, where $\mathbf{I}$ is the identity matrix.[20] Writing $\mathbf{X} = \mathbf{F}_0$ and $\widehat{\boldsymbol{\Sigma}} = \widehat{\Phi}_0$ for convenience, the process of our analysis would for each $j = 1, 2, \ldots r$ have established certain relations $\widehat{\Phi}_{j-1} = \widehat{Q}_j \widehat{\Phi}_j \widehat{Q}_j^{\tau} + \widehat{\Psi}_{j-1}$ now, where $\widehat{Q}_j$ estimates a coefficient matrix and $\widehat{\Psi}_{j-1}$ estimates the specific variance of $\mathbf{F}_{j-1}$. If $\widehat{\Phi}_r \approx \mathbf{I}$ very well, we may take $\widehat{\Phi}_{r-1} = \widehat{Q}_r \widehat{Q}_r^{\tau} + \widehat{\Psi}_{r-1}$ as our new estimate of Φ_{r-1}; set $\Lambda_{r-1} = [\,\widehat{Q}_r \,|\, \widehat{\Psi}_{r-1}^{\frac{1}{2}}\,]$ where $\widehat{\Psi}_{r-1}^{\frac{1}{2}}$ is the diagonal matrix whose entries are the square roots of the corresponding entries of $\widehat{\Psi}_{r-1}$; as $\widehat{\Psi}_{r-1}$ is diagonal we have $\widehat{\Phi}_{r-1} = \Lambda_{r-1} \cdot \Lambda_{r-1}^{\tau}$ then. Let $m_1 > m_2 > \ldots > m_r$ number the factors in the successively isolated ensembles; note that Λ_{r-1} is an $m_{r-1} \times (m_{r-1} + m_r)$ matrix. Writing Λ_{r-2} for $[\,\widehat{Q}_{r-1} \cdot \Lambda_{r-1} \,|\, \widehat{\Psi}_{r-2}^{\frac{1}{2}}\,]$ next, the relation $\widehat{\Phi}_{r-2} = \widehat{Q}_{r-1} \widehat{\Phi}_{r-1} \widehat{Q}_{r-1}^{\tau} + \widehat{\Psi}_{r-2}$ sanctions a re-estimating decomposition $\widehat{\Phi}_{r-2} = \Lambda_{r-2} \cdot \Lambda_{r-2}^{\tau}$ again, with Λ_{r-2} being an $m_{r-2} \times (m_{r-2} + m_{r-1} + m_r)$ matrix. Continuing thus we will arrive at a re-estimating decomposition $\widehat{\Phi}_1 = \Lambda_1 \cdot \Lambda_1^{\tau}$ of the correlations between the factors first isolated, with Λ_1 being an $m_1 \times (m_1 + m_2 + \ldots + m_r)$ matrix. One sets $L = \widehat{Q}_1 \cdot \Lambda_1$ customarily at this point, and the relation $\widehat{\boldsymbol{\Sigma}} = \widehat{Q}_1 \widehat{\Phi}_1 \widehat{Q}_1^{\tau} + \widehat{\Psi}_0$ sanctions the re-estimate $\widehat{\boldsymbol{\Sigma}} = L \cdot L^{\tau} + \widehat{\Psi}$ now, with $\widehat{\Psi} = \widehat{\boldsymbol{\Sigma}} - L \cdot L^{\tau} \approx \widehat{\Psi}_0$ being our re-estimate of the specific variance $\boldsymbol{\varepsilon}$ of $\mathbf{X}$. Here L is a $p \times (m_1 + m_2 + \ldots + m_r)$ matrix and, taking each of its columns to identify a distinct component of some ensemble $\mathbf{F}$ of *mutually independent* latent factors now, one usually takes L to estimate how the components of $\mathbf{F}$ would load the components of $\mathbf{X}$ were $\mathbf{F}$ to factorize $\mathbf{X}$; in which case L would also estimate the coefficient matrix in some equation $\mathbf{X} = \Lambda \mathbf{F} + \boldsymbol{\varepsilon}$ of course, since we would have $\boldsymbol{\Sigma} = \Lambda \Lambda^{\tau} + \Psi$ here.

Recalling how loading matrices are obtained from products of coefficients matrices and factor correlations, the procedure of Schmid and Leiman may be said to recast as loadings of $\mathbf{X}$ by $\mathbf{F}$ the loadings of factor ensembles by succeeding ensembles in the series $\mathbf{F}_1, \mathbf{F}_2, \ldots, \mathbf{F}_r$ of successive factorizations. The propriety of transforming many ensembles of intercorrelated latent factors into one ensemble of independent factors would depend, one supposes, on how the 'independence' of the factors themselves is reflected in the statistical independence of the random variables that their measures would be; but the matter does not seem to worry psychometricians overmuch.[21] There is one complication worth noting, though, and we shall bring that out presently. Putting aside the merits of orthogonalizing them, however, we note that some consensus seems to have emerged on how many levels of latent factors a satisfactory description of intelligence would require, and on how those levels and their constituent factors should be characterized.

---

20 We are assuming, recall, that all latent factors have unit variance.

21 In the paper which presented their procedure Schmid and Leiman do not address the matter; and the procedure is recommended to factor analysts on the grounds that "*an oblique factor solution often tends to confound the resulting interpretation*". (Cf. THE DEVELOPMENT OF HIERARCHICAL FACTOR SOLUTIONS, *Psychometrika, Volume 22, No.1, March 1957*).

Interested readers may acquaint themselves with these particulars from some authoritative text.[22] We return now to matters aired previously. Suppose $F_1, F_2, \dots, F_q$ are the finally imputed factors of intelligence, derived from manifest variables $X_1, X_2, \dots, X_p$ which numerize as scores individual performances on the p distinct tasks set by some putative test of intelligence, through a hierarchical factor analysis of such performances: these factors would describe the 'specific abilities' mentioned earlier. The jointly expected values $f_1, f_2, \dots, f_q$ of these factors can now be estimated for any jointly observed values $x_1, x_2, \dots, x_p$ of the variables: these would be examples of the 'tractable transformations of raw scores' talked of then. Were these final factors sufficiently correlated, the putative ability of 'general intelligence' could be supposed the latent factor to which their covariance and common variance is due: provided that these quantities are sufficiently conserved by one such factor. The final factors are indeed found to be sufficiently correlated in factor analyses performed on the scores obtained on many intelligence tests, and a latent factor of these final ones often conserves their covariance and common variance sufficiently. The putative general intelligence so identified has come to be called the g factor; g may be estimated in a number of ways, each of which amounts to a 'weighting and summing' of some tractable transformation of scores.[23] The complication attending the procedure of Schmid and Leiman adverted to above emerges here. After that orthogonalization g will be the last in a series $F_1, F_2, \dots, F_m$ of statistically independent factors. But then the expected value of our measure of g would equal the conditional expected value $E(g \mid f_1, f_2, \dots, f_{m-1})$ for *any* particular measured values of $F_1, F_2, \dots, F_{m-1}$ at all: and that a measure of general intelligence should behave so seems very odd, however these less than general factors may be characterized. But this is an entirely evident difficulty of course, mathematically considered, and numerate psychometricians will no doubt have already resolved it to their satisfaction.

---

[22] A convenient reference here is the aforenoted INTELLIGENCE: A VERY SHORT INTRODUCTION by Ian Deary; James Carroll's HUMAN COGNITIVE ABILITY is the most cited one.

[23] An estimate of the *first* principal component of $\mathbf{F} = (F_1, F_2, \dots, F_q)^\tau$ is often used to identify g; recall that this is $\mathbf{u}^\tau \cdot \mathbf{F}$ where $\mathbf{u}$ is the eigenvector associated to the largest eigenvalue of the covariance Φ of $\mathbf{F}$. Geometrically considered $\mathbf{u}^\tau \cdot \mathbf{F}$ is the variable obtained by projecting $\mathbf{F}$ on the axis determined by $\mathbf{u}$ in $\mathbf{R}^q$: which best captures total variance when a variable vector is so projected, and projection on this axis captures total variance to the extent that the largest eigenvalue exceeds all the others. The eigenvector determining the first principal component is estimated using the estimate of Φ got in estimating the factorization $\mathbf{X} = W\mathbf{F} + \delta$ of $\mathbf{X}$ by this final $\mathbf{F}$; and the estimated value $\mathbf{f}$ of $\mathbf{F}$ weighted by the components of this vector estimates g. Another method of weighting these estimated factor values estimates the specific variance Ψ_f of $\mathbf{F}$ itself, and takes for weights the components of an estimate of the eigenvector corresponding to the largest eigenvalue of $\Phi - \Psi_f$; this is called the method of *the first principal factor*. Let $\mathbf{u}_f$ denote either of these estimated vectors; then $(\mathbf{u}_f)^\tau \cdot \mathbf{f}$ is the estimate of g. Let $\mathbf{v}_f$ be the vector got by scaling $\mathbf{u}_f$ with the square root of its associated eigenvalue. Yet another way of estimating g, consistent with how the factorization of $\mathbf{X}$ by factors of $\mathbf{F}$ was modeled above, would use values $\mathbf{x}$ of $\mathbf{X}$ itself and solve $\mathbf{x} = \mathbf{g}\,(\widehat{W} \cdot \mathbf{v}_f) + \widehat{\Psi}_f$ by weighted least squares regression again, where $\widehat{W}$ and $\widehat{\Psi}_f$ estimate the coefficients W and the specific variance Ψ_f respectively. After the orthogonalization of Schmid and Leiman the transform of the single final factor identifies g, and the corresponding factor score estimates it.

Let us finally consider how empirical any measure of g could be, given that one goes beyond the data themselves in the course of imputing latent factors. One way to proceed here might be to estimate the aforenoted first principal component of the scores on the tasks $\mathbf{X}$ themselves: and to regard a measure of g as empirical to the extent that its values are positively correlated with the latter values. When the measure of g correlates suitably with the first principal component of $\mathbf{X}$, as it often will, one might ask if the latter and more empirical measure makes g superfluous. But by characterizing general intelligence as it does, through suitable descriptions of the latent factors it weights, g may well provide some grounds for believing that its summary values are adequately measuring whatever it is that common uses of the word "intelligence" refer to. The large reason why is that different individuals are likely to have acquainted themselves in very different ways with the latent features of their world. To communicate their understandings of such features to one another individuals must align their differing experiences somehow, let me venture to say, and discover or devise shareable aspects of latent features. The latent factors isolated by factor analysis provide such aspects, one is tempted to say: whose shareability depends on the descriptions, in a daily language now, that these factors admit. Whether factor analysis as it is currently done captures how alignments of individual experience to shareable aspects are actually effected is another matter, of course; my claim requires only that the *results* of factor analysis provide shareable aspects of latent features. The instances which have separately intimated intelligence to different individuals are likely to have varied considerably between themselves: simply because intelligence can be exhibited in great many things that human beings do. The tasks found on any given test of intelligence would reflect its makers' experience of intelligent behaviour and, however many or clever these makers may be, they are not likely to have enjoyed a thorough acquaintance with the manifold instancing of intelligence: unless they are, between themselves, acquainted with almost everything that requires intelligence. The determination of scrutable and broadly specified latent factors should rescue a summary measure like g, therefore, from the inevitably parochial experience of intelligent behaviour that the makers of any test will have had: provided, of course, that the final factors isolated from different tests admit conformable descriptions in some natural language shared by factor analysts of intelligence.[24]

The positing of g happened very early in psychometry, long before quantitative techniques had firmed up; and occasioned a deal of polemic actually, with some psychometricians regarding it as a statistical artefact.[25] But the case for the 'reality' of g gained considerable ground once its measures began to corre-

[24] Factors of intelligence have been constructed using data from a number of tests. James Carroll collated these factorizations in HUMAN COGNITIVE ABILITY; the congruence of different ensembles of final factors is a matter he must have considered.

[25] Charles Spearman had postulated and devised measures of general intelligence quite soon after intelligence tests were first devised; and factor analysis seems to have had its beginning there. The widely reported slanging match between Stephen Jay Gould and the votaries of intelligence testing went on around g. Unluckily for Gould his characterization of factor analysis was somewhat careless, and his criticisms have been scanted on that account.

late in expected ways with certain measures of the brain itself.[26] One notable use of g is to distinguish those tasks on an intelligence test which might be testing specific abilities that are not generally shared: in which case g should not too much load the variables which record scores on those tasks.[27] To save writing we shall now talk of g loading the tasks themselves. The unexpected rise over the years in the mean scores achieved on intelligence tests has been explained away by showing that the gain is related to the later inclusion of relatively more tasks which g does not load very much. More controversially, the case for racial differences in intelligence has come to depend on showing that the average performance on tasks which g loads highly varies significantly between racial groups.[28] The postulation of g remains a contentious matter, however, and interested readers should find copious reference on the Web to its uses and abuses. Our summary account of factor analysis concludes here, with the hope that the mathematically averse have not been made more so by it; but we shall have to refer again to the detail of the procedure, as we consider again the question about meaning which had concluded the previous section.

3

We had supposed, earlier, that the meaning of a word which names a putative latent feature of the world consists in some socially prevalent grouping and differential weighting of the manifest features of those objects or events which are thought to possess that feature; and the claim to be examined then was that factor analysis mirrors how prevalent groupings and weightings of manifest features come about. But where analysis discloses a hierarchy of latent factors one might expect a congruent layering of meaning; and the claim must

[26] Ian Deary's recent LOOKING DOWN ON HUMAN INTELLIGENCE explores this.

[27] Let D be the diagonal matrix with $(D)_{jj} = (\widehat{\sigma}_{jj})^{-1}$ again, where $\widehat{\sigma}_{jj}$ estimates the variance of X_j for $j = 1, 2, \ldots, p$. Using the notation of the previous footnote, when the vector $\widehat{W} \cdot \mathbf{v}_f$ identifies g, the j-th component of the vector $D \cdot \widehat{W} \cdot \mathbf{v}_f$ estimates how it loads X_j. When g is identified by some weighting $\sum_k w_k F_k$ of the final factors, how it might load any X_j is estimated by the correspondingly weighted sum of the estimated loadings of X_j by the $\{F_k\}$. With the method of Schmid and Leiman, which seems most used in practice, the loadings of the $\{X_j\}$ by g may be read off the appropriate column in the final matrix of loadings.

[28] The question of whether 'blacks' are less intelligent then 'whites' seems to exercise North Americans no end: even though there are any number of seemingly intelligent black people and any number of seemingly retarded white people, and despite the difficulties one would face in *explaining* how pigmentation could affect intelligence. Some 'genetic' account would have to be proferred now, and one would either have to specify how the genetic structures which determine pigmentation could affect the structures which determine intelligence, or specify a 'black genome' which, considered as a whole, happens to compromise the latter genetic structure; and if one managed either of these feats, the circumstance that intelligent blacks are not extremely rare would be a great mystery. The claim that "blacks are less intelligent than whites" is only exiguously one about groups, it might now be urged, and asserts only that the average of the scores achieved by individual blacks on g-loaded tasks in tests of intelligence is less than the average of the scores achieved there by individual whites; but note then that the word "intelligent" in the general claim would *mean* something other than what it does in the sentence "Condoleeza is less intelligent than George."

now be that factor analysis mirrors the process which determines the prevalent groupings and weightings of subsidiary features, either manifest or latent, of objects or events thought to possess a latent feature. Were that indeed so, and provided the 'intuitions' which guide pyschometricians as they devise tests of intelligence are not peculiar to their tribe, the factors successively identified by hierarchical factor analyses of performances on such tests would make explicit common recognitions of intelligence, let us say, which are implicit in the way intelligence is commonly assessed in the ordinary course of life; and the final or g factor would disclose how manifestations of intelligence are commonly grouped and weighted when intelligence is ordinarily assessed. An unexamined assumption here is that the daily meaning of "intelligence" is shaped most by the comparative assessment of individual intelligence.

Suppose that ℓ and ℓ' are words naming distinct latent features, and suppose φ is some subsidiary feature, manifest or latent, that is common to objects or events manifesting whatever ℓ and ℓ' name. Take the words "brave" and "rash" as examples here, and suppose the common subsidiary feature is *standing fast* when it would be safer to withdraw. In deciding whether someone standing fast in such a situation were being brave or rash one would certainly group and weigh that standing fast with and against other features of the situation; and it seems true enough that the difference between bravery and rashness would lie in how such steadfastness is of a piece, one might say, with whatever else is done. Now suppose some process analogous to factor analysis has determined the meanings of these words by working upon instances of bravery and rashness; the process would have worked upon such instances *considered together rather than separately,* one thinks, simply because it is often difficult to distinguish between bravery and rashness. What our example advertises may be summed up thus: distinct latent features that share determining subsidiary features are unlikely to be understood without reference to each other. The effective redaction of individual understandings into common meaning depends, one suspects, on conspicuously shared subsidiary features actively involving with each other the meanings of words that name latent features.

Returning to the business at hand, consider the recent fortunes of the word "creative" : which is often contraposed to "intelligent" now in ways our literate parents would have found odd.[29] The words "creative" and "creativity" seem to serve nowadays much as "imaginative" and "imagination" once did; and I shall use the latter in their stead. The consensus psychometry seems to have reached on the final factors of intelligence finds these divided, among other kinds, into *fluid* and *crystalline* ones; improvisation of some sort seems to distinguish the former from the latter, and if the factor analysis has indeed made explicit something implicit in common uses of the word "intelligence", then the fluid or improvisatory character of intelligence cannot be anything remote from everyday experience. But imagination would have to improvise well, at the very least, whatever it is commonly taken to be; and so imagination would have to be a

[29] A potent example is needed; I hope this will serve. Auden may fairly be called a clever poet — cleverer than Yeats, for example, and a lesser one for just that reason perhaps — but his 'creativity' was surely not a thing apart from the intelligence that his cleverness reveals.

fluid thing as well. Whether or not intelligence is commonly regarded as linked to imagination through a shared fluidity would be hard to determine, but the question need not detain us: the linkage may be implicit. How closely their fluid character links intelligence and imagination is anyone's guess now: and one might then wonder how much about intelligence is masked by a regime of testing which, regarding intelligence and imagination as independent or contrasting powers of mind, does not specially assay imagination along with intelligence.[30] So, while the factor analysis of scores achieved on current tests of intelligence might make explicit *some* of what is implicit in the common understanding of intelligence, one cannot say how much about intelligence — as that is commonly understood or as it may in fact be — is thereby obscured; and now one might ask whether or not current modes of testing *adequately* measure intelligence.

Let us next consider, in a more general way, whether or not factor analyis could plausibly be said to mirror how words naming latent features acquire their meaning. The attraction of the claim must lie in the conserving, within ensembles of successively extracted factors, of the covariance and common variance of a suite of underlying variables. The 'efficiency' of natural languages suggests that some analogous process of consolidation shapes meaning there: words naming latent features could not be used flexibly otherwise, one thinks, and natural languages would have to effect such a consolidation somehow. The immediate complication for our claim, however, is the presupposition that such a word gets underway in common usage only after manifold instancings of the putative latent feature it has come to name: which instancings would have established, already, correlations between observable features manifesting this latent one. On this view words naming latent features seem passive carriers of senses consolidated and lying implict in general use before their advent. But many such words might have, rather, gained their common uses by actively colonizing remote or unclaimed tracts of experience. Such 'colonizations' are usually begun with compelling tropes, I shall risk saying, and completed when what began as figured speech becomes literal: 'dead' metaphors are those that have gone native, one might say. Now when words naming latent features come to their common senses so they will direct, one thinks, the pertinent manifesting of what they name. As their meanings sharpen they would determine how what they name will be given to experience: in which case the 'data' would presuppose the meanings that the semantic analogue of factor analysis ostensibly produces from them.

But let us suppose, for the moment, that the meanings of words naming latent features derive from prior correlations between manifest features, which correlations they have played no part in determining. Suppose ℓ and ℓ' name distinct latent features, and suppose X and X' are features which each manifest both ℓ and ℓ'. On the view we are considering there is some absolute correlation between X and X' which determines how they figure, along with whatever else manifests ℓ and ℓ', in the distinct groupings and

[30] The 'objective' assay of imaginative power would very likely require protocols entirely unlike those which currently regulate the assessment of intelligence; and so may present methodological obstacles that psychometry, as that is currently practiced, cannot easily surmount.

weightings of manifest features that the distinct meanings of ℓ and ℓ' consist in. That X and X' could correlate differently when they manifest the distinct denotata of ℓ and ℓ' seems the more natural assumption, but the difficulty there for our claim is plain: differing correlations could be established as such, as correlations pertinent to one or other latent feature, only if whatever ℓ and ℓ' name are already understood as distinct latent features. Now if factor analysis as currently practiced is to mirror the formation of meaning, much would depend on the styles of rotation that have become established in the assay of latent features; and, as with the assay of intelligence, the dominant style of rotation seems to seek out those univocal subsidiary features that perfect indicators would be. The corollary condition imposed on natural languages thereby would be that observable features of the world are selected as such from compresent ones, by common usage, in such a way that each indexes only one of the ensemble of latent features that a given array of observables might manifest. Whether or not common usage generally works so, and in a world amenable to such operations, would be very hard to say.[31] One could wish for some 'transcendental' consideration, such as Wittgenstein might have divined, that a natural language simply could not work so: but none lies ready to hand. Common usage does not seem to cull 'appearances' so in the natural language that English currently is: and the smile that would greet a psychometrician who really supposed otherwise might well display, even to such an innocent, some satisfaction accompanying the evident amusement. But this might be happenstance merely; and perhaps one should try to imagine circumstances which would render common usage a rigid censor of phenomena, registering in its lexicon only such as unequivocally announce their parent noumena.[32]

[31] What common usage is taken to work against here is a local sort of polysemy: which, using the language of Saussure, one might term the compresence of distinct 'signifieds' for a given 'signifier' in the phenomenal neighbourhood, as it were, of a latent feature. The 'sign' that a latent feature would engender in a language so constrained might initially be diagrammed as a tree, whose leaves consist of observable signifiers, grouped into clusters whose members each lead back to one node, with the subsidiary signifieds that these nodes are grouping themselves into clusters that each lead back to one node, and so on: until we reach, at its root, the signified that serves as the concept of the latent feature. In the picture we have so far observable signifiers and subsidiary signifieds are arrayed at descending levels, with no branches between levels that are not adjacent. The branches connecting a signified at some level to the signifieds or signifiers clustered immediately above it may be imagined thick in proportion to how the latent feature conceptualized there loads those features which it lay latent in. What should complicate the picture is the loading of variables by factors of factors: the semantic analogue of which will introduce branches that cut across intervening levels. Even so this tree of meaning would have pleased a logical positivist, with its serried leaves relaying in concert some world's lucid intimations; and would be dismissed as naive by philosophers for whom words consort in Byzantine ways, along Quinean webs say, whose weave is rarely strained by what the senses net.

[32] Had English polity become an 'enterprise association' early enough, dedicated to the building of Bacon's Bensalem say, common usage might well have enforced generally the demand for plain speech that the Royal Society's founding members placed on each other: and signs that were frank tokens of wonders would rank highly, one thinks, among the various 'efficiencies' prized by persons engaged with each other just so. The phrase "enterprise association" is Micheal Oakeshott's: and the social mode so named is distinguished from 'civil association' by some *telos,* formally acknowledged as such by those joined so, to the achieve-

Such exercises require rather more leisure than we have, though, and I must turn now to an assumption granted without demur thus far, namely, that the practice of pyschometry makes no difference to how the word "intelligence" is used after the business of measuring intelligence gets underway. The current and common contraposition of "creative" to "intelligent" suggests otherwise; and, given the social uses to which intelligence tests have been put in America and England, one may hazard supposing that the fact of testing has appreciably affected how the 'testees' themselves come to use such words. Assuming, as we have been, that the common use of "intelligence" is determined by the comparative assessment of intelligence, one should ask how the assessment individuals make of each other's intelligence on the basis of their scores on a test would conform to what assessments they would otherwise have made. The score one receives on a test allows one to compare one's intelligence in *a uniform way* with every other testee's. Now one might wonder if even those disposed to compare themselves with everyone around them would have assayed the intelligence of their every acquaintance in the same way, had they the social license to do so even. Whether or not the practice of measurement has remade common uses of intelligence would depend, then, on the extent to which the comparison of scores on tests — however directly or indirectly effected, through social choices institutionalized on their basis say — has replaced earlier ways in which individuals assayed each other's intelligence: which would have varied, one thinks, according to their social distance from each other.

The question may be pursued by asking if scores on tests have come to be used *criterially,* rather than as *evidence* merely, in the comparative ascription of intelligence. The difference between criteria and evidence is not easily summarized; but I trust that the daily meanings of "criteria" and "evidence" will see me through here. Deciding the question generally would be difficult: one would, for instance, have to gauge the extent to which testees discount the institutional grading of their intelligence on the basis of such scores.[33] But if these scores have come to be criterially employed by testees, then the measuring of intelligence will certainly have affected how "intelligence" is commonly used. The grouping and weighting of manifest features that testees were once

ment of which the remaining terms of their association are subordinate. In his tract *On Human Conduct,* where their differences were explored, modern European polities were seen as uneasy mixtures of these divergent modes of association; and that 'enterprise' so often trumped the restraints of 'cives' would have been due, in no small way, to the technical advance upon prior material culture that industrial modernity made. The 'play of appearance' English allows traces the civil lineaments of those polities which the language animates, one is now tempted to say: and one might even hope that such play, indulged enough, will release their citizens from the thrall of the transfiguring *techne* that 'mechanical' modernity and its 'electronic' progeny have loosed upon them. Going on thus would be naive. Even so, just how the 'sciences of man' abetted that suborning of civil association by enterprise which technological prowess forced upon polities will emerge, one hopes, as a principal theme in the historiography of our era, when historians are remote enough from us to regard our doings coldly: provided the 'progress' of these sciences has not obliterated the distinction.

[33] Proportionally fewer Americans than Britons are likely to do so, one thinks; and that circumstance seems, obscurely, a concomitant of the 'manufacture of consent' in America which Chomsky has detailed.

allegedly able to do, however implicitly, will have given way to a reliance on summary numbers; and the process of measurement, had it indeed elicited the 'core meaning' that lay implicit in daily uses of the word, would have gone on to erase just that understanding from the public mind. One can only hope that the common understanding of intelligence that would now lie immured in its summary measure was an accurate one.

I should now point to some institutional practice which would use test scores more criterially than evidentially; and their employment in sorting recruits or conscripts to the armed forces provides the nearest example. As it happens, one of the first uses of intelligence testing on a large scale was to sort recruits to the infantry in the First World War: where their grading by putative intelligence seems to have decided just how expendable they would be in battle.[34] The testing of intelligence for military purposes is likely to be a more nuanced affair nowadays, of course, but the example displays just why psychometricians would want accurate and adequate measures of intelligence. Summary measures derived in some uniform way from standardized tests should insulate testing from the 'subjectivity' of recruiters; and, however valuable the experience of seasoned recruiters might be, the 'objectivity' of a uniform sorting according to intelligence seems an index of some superior 'fairness' as well, given the consequences that will attend the sorting. Our example also suggests why, in democratic polities whose citizens make a fetish of it, any summary measure of intelligence would have to be delivered by an empirical model: models that were hostage to prior assumptions would violate, one thinks, the particular sense of equity that distinguishes the democratic from other sorts of political animal. The factor analytic construction we have reviewed begins with evidence of intelligence that would be received as such by everyone, presumably, and in working up a summary measure makes no special assumptions about intelligence. What finally makes the proceeding 'scientific' is the adoption of a particular method of rotation: which ensures that results are replicable. The method of rotation that happens to have been adopted forces poor assumptions upon us only when factor analysis is taken to mirror the formation of meaning in a natural language; and we must now consider how central this hypothesis may be to what we have called the mathesis of intelligence.

The factor analysis of data from a number of different tests of intelligence appears, in fact, to have yielded conformable ensembles of factors arrayed in congruent hierarchies: and so provided a taxonomy of intelligence that psychometricians have largely come to agree upon.[35] Our hypothesis warrants saying that this taxonomy is implicit in the common understanding of intelligence. But why this circumstance should recommend the taxonomy is very hard to see; after all, common understanding may be as mistaken here as it apparently was

[34] The morality of selectively endangering the less intelligent is a simpler matter in war, one can only suppose, than it would otherwise be.

[35] We mentioned Carroll's HUMAN COGNITIVE ABILITY with regard to this. That psychometry has only provided such taxonomies thus far is Deary's view in LOOKING DOWN ON HUMAN INTELLIGENCE.

about the taxa of sea creatures.[36] The claim does have a polemical use which we should mention. Some critics of intelligence testing have maintained that the summary numbers produced from tests could not measure what intelligence is ordinarily taken to be; the hypothesized link of method to daily meaning affords a very short way with that objection. But, again, it is not clear why psychometry should be bound by ordinary understandings of mind: especially if we regard brains as embodying minds. We could not, obviously, regard the brain so unless we already had some secure purchase on the mental. But that ordinary understandings of mind always afford such purchase is far from clear: consider, for instance, how the physiology of colour has come to be understood.[37] All told, there is little reason why a science of the mental should observe the proprieties that natural languages demand: and as a way of world making, in Nelson Goodman's felicitous phrase, a developed science should probably be immune to how its terms are commonly understood.[38] What complicates matters for the social and behavioural sciences, of course, is the particular social uses to which they are put.[39] These seem to have developed more as modes of *praxis* than *theoria* generally: and perhaps inevitably so.[40] The social process that the quantification of these sciences has become, in America particularly, should reward study. If the prospect of gains in 'objectivity' is what sustains their practitioners' drive to quantify, what bears thinking on is the extent to which a distinctive numerization of objectivity has been forced by particular exigencies: by a style of democracy that, in order to preserve its 'way of life' allegedly, has had to abet an imperial exercise of power through modes of force conceived and applied in oblique ways.[41]

I would have liked to end by considering what the study of intelligence might have been had the business of measurement been an ancillary affair. We might have speculated, for instance, on what historiography would have to be for a

[36] That common understanding is zoologically mistaken maybe of little consequence to common understanding, however, and I do not mean to deprecate it. The man who nets dolphins while fishing for tuna, for instance, is no less a fisherman on that account.

[37] A good text here is COLOUR FOR PHILOSOPHERS by C.L.Hardin (*Hackett, 1988*).

[38] Consider that physics neither constrains, nor is specially constrained by, how such terms of art as "force" and "mass" are elsewhere used.

[39] Recall the procedure of Schmid and Leiman, which is intended to circumvent the 'confounding' by each other that the measures of oblique factors woud undergo. But such confounding would be inconvenient only when these measures are used in very specific ways: to sort recruits to the military according to their suitability for specific combat roles, for instance, in a methodical and readily repeated way. The distinctive operabilities of complex weapons would make their use a far more important consideration now than ever before, one thinks, in the testing of intelligence for military purposes.

[40] The social and behavioural sciences are modes of praxis to the extent that their application remakes, continually, the social ground upon which they build. Gaston Granger conceived these sciences as modes of praxis precisely, in his magisterial FORMAL THOUGHT AND THE SCIENCES OF MAN, and was prepared to celebrate them were they properly conducted as such.

[41] Consider how the sanctions on Iraq between the first Gulf War and the second were 'legally' enforced; and I am groping, with the word "oblique", toward how the progress of weaponization seems to have transformed military violence. One wonders, for instance, how 'costs and benefits' were assessed in deciding to the pack with depleted uranium the heavier munitions used in Iraq; or, going back to an earlier American adventure, how the 'doses' were determined in deciding to defoliate with dioxin large tracts of the Vietnamese countryside.

history of ideas to specially acquaint us with human intelligence, in ways admitting that 'abduction' of potent generalizations which distinguishes a developed science. But such exercises would seem only fanciful to both historians and psychometricians, one suspects, were they comprehensible at all. We had noted how measures of the brain are being brought to bear on measures of intelligence; the mathesis of intelligence will surely mutate as the physical assaying of the brain becomes more and more sophisticated, and thinking on that seems more to the point. One can only hope that technical sophistication will be matched by a refinement in conceiving.[42]

[42] Considered as a social process the measurement of intelligence seems a good example of Baconian science: a collective undertaking that "levels men's wits" in the course of methodically producing some useful thing. That the measuring of intelligence should proceed so is not an irony pyschometricians would dwell on, one supsects; and there would be little use now in repeating for their hapless subjects Auden's counsel to *"not sit with statisticians, nor commit a social science"*.