THE FACTORIAL ANALYSIS OF HUMAN ABILITY

By

SIR GODFREY THOMSON

D.C.L., D.Sc., Ph.D.

University of Edinburgh

UNIVERSITY OF LONDON PRESS LTD.

WARWICK SQUARE, LONDON, E.C.4

First Printed	. . . . .	1939
Second Edition	. . . . .	1946
Third Edition	. . . . .	1948
Fourth Edition	. . . . .	1950
Fifth Edition	. . . . .	1951

Printed & Bound in England for the University of London Press Ltd.,
by Hazell, Watson & Viney, Ltd., Aylesbury and London

PREFACE TO THE FIRST EDITION

THE theory of factorial analysis is mathematical in nature, but this book has been written so that it can, it is hoped, be read by those who have no mathematics beyond the usual secondary school knowledge. Readers are, however, urged to repeat some at least of the arithmetical calculations for themselves.

It is probable that the subject-matter of this book may seem to teachers and administrators to be far removed from contact with the actual work of schools. I would like therefore to explain that the incentive to the study of factorial analysis comes in my case very largely from the practical desire to improve the selection of children for higher education. When I was thirteen years of age and finishing an elementary school education, I won a " scholarship " to a secondary school in the neighbouring town, one of the early precursors of the present-day " free places " in England. I have ever since then been greatly impressed by the influence that event has had on my life, and have spent a great deal of time in endeavouring to improve the methods of selecting pupils at that stage and in lessening the part played by chance. It was inevitable that I should be led to inquire into the use of intelligence tests for this purpose, and inevitable in due course that the possibilities of factorial analysis should also come under consideration. It seemed to me that before any practical use could be made of factorial analysis a very thoroughgoing examination of its mathematical foundations was necessary. The present book is my attempt at this. . . . It may seem remote from school problems. But much mathematical study and many calculations have to precede every improvement in engineering, and it will not be otherwise in the future with the social as well as with the physical sciences.

GODFREY H. THOMSON

MORAY HOUSE,
UNIVERSITY OF EDINBURGH,
November 1938

PREFACE TO THE FIFTH EDITION

In earlier editions since the first, the chief changes in the second edition were that the original chapter on Simple Structure was expanded into three, to cover oblique factors and second-order factors, while Dr. D. N. Lawley provided a chapter on factor analysis by maximum likelihood, and a corresponding section in the mathematical appendix. The main changes in the third edition concerned the identity of simple structure factors after univariate selection, and the relations between two sets of variates. In the fourth, the principal addition was of Lawley's formulæ for the standard errors of individual residuals, and of factor loadings, when maximum likelihood methods have been used.

In the present (the fifth) edition it has for the first time been possible to reset the whole book. This has permitted more extensive alterations to be made, and the opportunity has been taken of rearranging the order of the chapters and recasting several of them, as well as inserting in their proper places in the text those pages which in former editions had to be added as appendices. Chapters V, VIII, and X will supply the minimum of technique, and the remainder of Parts II and III will give in addition a description of the methods of analysis using principal components, using the principle of maximum likelihood, or using Thurstone's Simple Structure.

I hope, however, that readers will not merely use the book as a set of recipes on how to carry out certain computations, but will study the geometrical explanations (twelve new diagrams have been added): and especially that they will ponder the implications of the two chapters, XVIII and XIX, on the influence of selection on factors, and the final two chapters on the sampling theory and certain fundamental questions.

GODFREY H. THOMSON

UNIVERSITY OF EDINBURGH,
April 1951

CONTENTS

CHAPTER IV. THE DEFINITION OF g

PART II. MULTIPLE-FACTOR ANALYSIS

CHAPTER V. THE CENTROID METHOD

CHAPTER VI. THE GEOMETRICAL PICTURE

Chapter VII. Principal Components

Chapter VIII. Testing Residues for Significance

Chapter IX. The Maximum Likelihood Method of Estimating Factor Loadings

PART III. THE ROTATION OF FACTORS

Chapter X. The Rotation of Factors

Chapter XI. Orthogonal Simple Structure

Chapter XII. Oblique Factors

Chapter XIII. Second-order Factors

PART IV. THE ESTIMATION OF FACTORS

CHAPTER XIV. REGRESSION AND MULTIPLE CORRELATION

CHAPTER XV. THE ESTIMATION OF A MAN'S FACTORS

PART V. CORRELATIONS BETWEEN PERSONS

CHAPTER XVI. REVERSING THE RÔLES

PART VII. THE NATURE OF FACTORS

CHAPTER XX. THE SAMPLING THEORY

CHAPTER XXI. SOME FUNDAMENTAL QUESTIONS

LIST OF DIAGRAMS

All science starts with hypotheses—in other words, with assumptions that are unproved, while they may be, and often are, erroneous ; but which are better than nothing to the searcher after order in the maze of phenomena.

T. H. Huxley

I am not insensible of the advantage which accrues to Applied Mathematics from the co-operation of the Pure Mathematician, and this co-operation is not infrequently called forth by the very imperfections of writers on Applied Mathematics.

R. A. Fisher

PART I

THE TWO-FACTOR THEORY AND ITS EXTENSIONS

CHAPTER I

THE THEORY OF TWO FACTORS

1. *Factor tests.*—The object of this book is to give some account of the "factorial analysis" of ability, as it is called. In actual practice at the present day this science is endeavouring (with what hope of success is a matter of keen controversy) to arrive at an analysis of mind based on the mathematical treatment of experimental data obtained from tests of intelligence and of other qualities, and to improve vocational and scholastic advice and prediction by making use of this analysis in individual cases. It is a development of the "testing" movement—the movement in which experimenters endeavour to devise tests of intelligence and other qualities in the hope of sorting mankind, and especially children, into different categories for various practical purposes; educational (as in directing children into the school courses for which they are best suited); administrative (as in deciding that some persons are so weak-minded as to need lifelong institutional care); or vocational, etc.

There are many psychologists who would deny that from the scores in such tests, or indeed from any analysis, we can (ever) return to a full picture of the individual; and without entering into any discussion of the fundamental controversy which this denial reveals, everyone who has had anything to do with tests will readily agree that this is certainly so at present in practice. But the tester may be allowed to try to make his modest diagram of the individual better, more useful, and if possible simpler.

Now, the broadest fact about the results of "tests" of all sorts, when a large number of them is given to a large number of people, is that every individual and every test is different from every other, and yet that there are certain rather vague similarities which run through groups of people or groups of tests, not very well marked off from one another but merging imperceptibly into neighbouring groups at their margins. To describe an individual accurately and completely one would have to administer to him all the thousand and one tests which have been or may be devised, and record his score in each, an impossible plan to carry out, and an unwieldy record to use even if obtained. Both practical necessity and the desire for theoretical simplification lead one to seek for a few tests which will describe the individual with sufficient accuracy, and possibly with complete accuracy if the right tests can be found. If, as has been said, there is some tendency for the tests to fall into groups, perhaps one test from each group may suffice. Such a set of tests might then be said to measure the " factors " of the mind.

2. *Fictitious factors.*—Actually the progress of the " factorial " movement has been rather different, and the factors are not real but as it were fictitious tests which represent certain aspects of the whole mind. But conceivably it might have taken the more concrete form. In that case the " factor tests " finally decided upon (by whom, the reader will ask, and when " finally " ?) would be a set of standards which, like any other standards, would have to be kept inviolate, and unchanged except at rare intervals and for good reasons. Some tendency towards this there has been. The Binet scale of tests is almost an international standard, and there is a general agreement that it must not be changed except by certain people upon whose shoulders Binet's mantle has fallen, and only seldom and as little as possible even by them. But the Binet scale is a very complex entity, and rather represents many groups of tests than any one test. By " factor tests " one would more naturally mean tests of a " pure " nature, differing widely from one another so as to cover the whole personality adequately. And since actual tests always are more or less mixed, it is understandable why " factors " have come to be fictitious, not real, tests, to be each approximated to by various combinations of real tests so weighted that their unwanted aspects tend to cancel out, and their desired aspects to reinforce one another, the team approximating to a measure of the pure " factor."

But how, the reader will ask, do we know a "pure" factor, how are we to tell when the actual tests approximate to it? To give a preliminary answer to that question we must go back to the pioneer work of Professor Charles Spearman in the early years of this century (Spearman, 1904). The main idea which still, rightly or wrongly, dominates factorial analysis was enunciated then by him, and practically all that has been done since has been either inspired or provoked by his writings. His discovery was that the "coefficients of correlation" between tests tend to fall into "hierarchical order," and he saw that this could be explained by his famous "Theory of Two Factors." These technical terms we must now explain.

3. *Hierarchical order.*—A coefficient of correlation is a number which indicates the degree of resemblance between two sets of marks or scores. If a schoolmaster, for example, gives two examination papers to his class, say (1) in arithmetic and (2) in grammar, he will have two marks for every boy in the class. If the two sets of marks are identical the correlation is perfect, and the correlation coefficient, denoted by the symbol r_{12}, is said to be $+ 1$. If by some curious chance the one list of marks is exactly like the other one upside down (the best boy at arithmetic being worst at grammar, and so on), the correlation is still perfect, but negative, and $r_{12} = - 1$. If there is absolutely no resemblance between the two lists, $r_{12} = 0$. If there is a strong resemblance, but falling short of identity, r_{12} may equal ·9; and so on. There is a method (the Bravais-Pearson) of calculating such coefficients, given the list of marks.* "Tests" can obviously be correlated just like

* The "product-moment formula" is—

$$r_{12} = \frac{\text{sum } (x_1 x_2)}{\sqrt{\{\text{sum } (x_1^2) \times \text{sum } (x_2^2)}}$$

where x_1 and x_2 are the scores in the two tests, measured from the average (so that approximately half the scores are negative), and the sums are over the persons to whom the scores apply. The quantity—

$$\sigma_1^2 = \frac{\text{sum } (x_1^2)}{\text{number of persons}}$$

is called the *variance* of Test 1, and σ_1 its *standard deviation.* If the scores in each test are not only measured from their average, but examinations, and a convenient form in which to write down the intercorrelations of a number of tests is in a square chequer board with the names of the tests (say *a*, *b*, *c* . .) written along the two margins, thus :

	a	*b*	*c*	*d*	*e*	*f*
a	.	·48	·24	·54	·42	·30
b	·48	.	·32	·72	·56	·40
c	·24	·32	.	·36	·28	·20
d	·54	·72	·36	.	·63	·45
e	·42	·56	·28	·63	.	·35
f	·30	·40	·20	·45	·35	.
Totals	1·98	2·48	1·40	2·70	2·24	1·70

It was early found that such correlations tend to be positive, and it is of some interest to see which of a number of tests correlates most with the others. This can be found by adding up the columns of the chequer board, when we see in the above example that the column referring to Test *d* has the highest total (2·70). The tests can then be rearranged and numbered in the order of these totals, thus :

		1 *d*	2 *b*	3 *e*	4 *a*	5 *f*	6 *c*
1	*d*	.	·72	·63	·54	·45	·36
2	*b*	·72	.	·56	·48	·40	·32
3	*e*	·63	·56	.	·42	·35	·28
4	*a*	·54	·48	·42	.	·30	·24
5	*f*	·45	·40	·35	·30	.	·20
6	*c*	·36	·32	·28	·24	·20	.

After the tests have been thus arranged, the tendency which Professor Spearman was the first to notice, and which

are then divided through by their standard deviation, they are said to be *standardized*, and we represent them by z_1 and z_2. About two-thirds of them, then, lie between plus and minus one. With such scores Pearson's formula becomes—

$$r_{12} = \frac{\text{sum of the products } z_1 z_2}{\text{number of persons } p}$$

In theoretical work, an even larger unit is used, namely $\sigma\sqrt{p}$. With these units, the sum of the squares is unity, and the sum of the products is the correlation coefficient. The scores are then said to be *normalized*, but note that this does not mean distributed in a " normal " or Gaussian manner.

he called "hierarchical order," is more easily seen. It is the tendency for the coefficients in any two columns to have a constant ratio throughout the column. Thus in our example, if we fix our attention on Columns *a* and *f*, say, they run (omitting the coefficients which have no partners) thus:

·54	·45
·48	·40
·42	·35
.	.
.	.
·24	·20

and every number on the right is five-sixths of its partner on the left.

Our example is a fictitious one, and the tendency to hierarchical order in it has been made perfect in order to emphasize the point. It must not be supposed that the tendency is as clear in actual experimental data. Indeed, at the time there were some who denied altogether the existence of any such tendency in actual data. Those who did so were, however, mistaken, although the tendency is not as strong as Professor Spearman would seem originally to have thought (Spearman and Hart, 1912). The following is a small portion of an actual table of correlation coefficients* from those days (Brown, 1910, 309). (Complete tables must, of course, include many more tests; in recent work as many as 57 in one table.)

	1	2	3	4	5	6
1	.	·78	·45	·27	·59	·30
2	·78	.	·48	·28	·51	·24
3	·45	·48	.	·52	·40	·38
4	·27	·28	·52	.	·41	·38
5	·59	·51	·40	·41	.	·13
6	·30	·24	·38	·38	·13	.

* In this, as in other instances where data for small examples are taken from experimental papers, neither criticism nor comment is in any way intended. Illustrations are restricted to few tests for economy of space and clearness of exposition, but in the experiments from which the data are taken many more tests are employed, and the purpose may be quite different from that of this book.

4. *G saturations.*—This tendency to "hierarchical order" was explained by Professor Spearman by the hypothesis that all the correlations were due to one "factor" only, present in every test, but present in largest amount in the test at the head of the hierarchy. This factor is his famous "*g*," to which he gave only this algebraic name to avoid making any suggestions as to its nature, although in some papers and in *The Abilities of Man* he permitted himself to surmise what that nature might be. Each test had also a second factor present in it (but not to be found elsewhere, except indeed in very similar varieties of the same test), whence the name, "Theory of Two Factors"—really one general factor, and innumerable second or specific factors.

It will be proved in the Mathematical Appendix* that this arrangement would actually give rise to "hierarchical order." Meanwhile this can at least be made plausible. For if Test *d* has that column of correlations (the first in our table) with the other tests solely because it is saturated with so-and-so much *g*; and if Test *b* has less *g* in it than *d* has, it seems likely enough that *b*'s column of correlations will all be smaller in that same proportion. We can, moreover, find what these "saturations" with *g* are. For on the theory, each of our six tests contains the factor *g*, and another part which has nothing to do with causing correlation. Moreover, the higher the test is in the hierarchical ranking, the more it is "saturated" with *g*. Imagine now a fictitious test which had no specific, a test for *g* and for nothing else, whose saturation with *g* is 100 per cent., or 1·0. This fictitious test would, of course, stand at the head of the hierarchy, above our six real tests, and its row of correlations with each of those tests (their "saturations") would each be larger than any other in the same column. What values would these saturations take ?

Before we answer this, let us direct our attention to the diagonal cells of the "matrix" of correlations (as it is called—a matrix is just a square or oblong set of numbers), cells which we have up to the present left blank. Since each number in our matrix represents the correlation of the two tests in whose column and row it stands, there should

* Para. 3 : and see also Chapter xviii, end of Section 6, page 283.

	g	1	2	3	4	5	6
g	1	r_{1g}	r_{2g}	r_{3g}	r_{4g}	r_{5g}	r_{6g}
1	r_{1g}	.	·72	·63	·54	·45	·36
2	r_{2g}	·72	.	·56	·48	·40	·32
3	r_{3g}	·63	·56	.	·42	·35	·28
4	r_{4g}	·54	·48	·42	.	·30	·24
5	r_{5g}	·45	·40	·35	·30	.	·20
6	r_{6g}	·36	·32	·28	·24	·20	.

be inserted in each diagonal cell the number *unity*, representing the correlation of a test with its own identical self. In these *self*-correlations, however, the specific factor of each test, of course, plays its part. These self-correlations of unity are the only correlations in the whole table in which specifics do play any part. These "unities," therefore, do not conform to the hierarchical rule of proportionality between the columns.

But the case is different with the fictitious test of pure g. It has no specific, and its self-correlation of unity should conform to the hierarchy. If, therefore, we call the "saturations" of the other tests r_{1g}, r_{2g}, r_{3g}, r_{4g}, r_{5g}, and r_{6g}, we see that we must have, as we come down the first two columns within the matrix—

$$\frac{r_{1g}}{1} = \frac{\cdot 72}{r_{2g}} = \frac{\cdot 63}{r_{3g}} = \frac{\cdot 54}{r_{4g}} = \frac{\cdot 45}{r_{5g}} = \frac{\cdot 36}{r_{6g}}$$

and similar equations for each other column with the g column, which together indicate that the six "saturations" are— ·9 ·8 ·7 ·6 ·5 ·4

Furthermore, each correlation in the table is the product of two of these saturations. Thus—

$$\cdot 72 = \cdot 9 \times \cdot 8$$
$$\cdot 42 = \cdot 7 \times \cdot 6$$
$$r_{34} = r_{3g} \times r_{4g}$$

The six tests can now be expressed in the form of equations :

$$z_1 = \cdot 9g + \cdot 436 s_1$$
$$z_2 = \cdot 8g + \cdot 600 s_2$$
$$z_3 = \cdot 7g + \cdot 714 s_3$$
$$z_4 = \cdot 6g + \cdot 800 s_4$$
$$z_5 = \cdot 5g + \cdot 866 s_5$$
$$z_6 = \cdot 4g + \cdot 917 s_6$$

Herein, each z represents the score of some person in the test indicated by the subscript, a score made up of that person's g and specific in the proportions indicated by the coefficients. The scores are supposed measured from the average of all persons, being reckoned plus if above the average and minus if below; and so too are the factors g and the specifics. And each of them, tests and factors, is "standardized," i.e. measured in such units that the sum of the squares of all the scores equals the number of persons. This is achieved by dividing the raw scores by the "standard deviation." The saturations of the specifics are such that the sum of the squares of both saturations comes in each test to unity, the whole variance of that test. Thus—

$$\cdot 436 = \sqrt{(1 - \cdot 9^2)}$$

5. *A weighted battery.*—This brief outline of the Theory of Two Factors must for the moment suffice. It is enough to enable the question to be answered which at the end of our Section 2 led to the digression. "How," the reader asked, "do we know a pure factor, how are we to tell when the actual tests approximate to it?" In the Two-factor Theory the important pure factor was g itself, and a test approximated to it the more, the higher it stood in the hierarchy. Its accuracy of measurement of g was indicated by its "saturation." And a battery of hierarchical tests could be weighted so as to have a combined saturation higher than that of any one member, each test for this purpose being weighted (as will be shown in Chapter XV) by a number proportional to $\frac{r_{ig}}{1 - r_{ig}{}^2}$, where r_{ig} is the g saturation of Test i (*Abilities*, p. xix). The battery saturation or multiple correlation with g is then—

$$\sqrt{\frac{S}{1 + S}}$$

$$\text{where } S = \Sigma \frac{r_{ig}{}^2}{1 - r_{ig}{}^2}$$

Although g remained a fiction, yet a complex test, made up of a weighted battery of tests which were hierarchical, could approach nearer and nearer to measuring it exactly, as more tests were added to the hierarchy. Each test added would have to conform to the rule of proportionality in its correlations with the pre-existing battery. If it did not do so it would have to be rejected. The battery at any stage would form a kind of definition of *g*, which it approached although never reached. And a man's weighted score in such a battery would be an estimate of *his* amount of *g*, his general intelligence. The factorial description of a man was at this period confined to one factor, since the specific factors were useless as description of any man. For one thing, they were innumerable; and for another, *being* specific, they were only able to indicate how the man would perform in the very tests in which, as a matter of fact, we knew exactly how he *had* performed.

6. *Oval diagrams.*—It is convenient at this point to introduce a diagrammatic illustration which will be useful in the less technical part of this book, although like all illustrations *it must be taken only as such, and the analogy must not be pushed too far.* If we represent the two abilities, which are measured by tests, by two overlapping ovals as in Figure 1, then the amount of the overlap can be made to represent the degree to which these tests are correlated. If we call the whole area of each oval the "variance" of that ability, we shall be introducing the reader to another technical term (of which a definition was given in the footnote to page 5). Here it need mean nothing more than the whole "amount" of the ability. The overlap we shall call the "*co*variance." If the two variances are each equal to unity, then the covariance *is* the correlation coefficient. To make the diagram quantitative, we can indicate in figures the contents of each part of

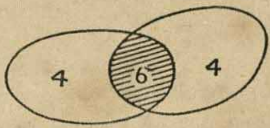


Figure 1.

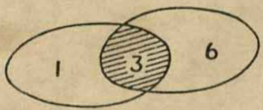


Figure 2.

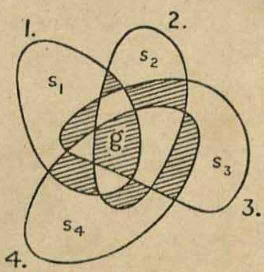


Figure 3.

the variance, as in the instance shown, which gives a correlation of $\frac{6}{10}$, or ·6. If the separate parts of each variance (i.e. of each oval) do not add up to the same quantity, but to v_1 and v_2, say, then the covariance (the amount in the overlap) must be divided by $\sqrt{v_1 v_2}$ in order to give the correlation. Thus, Figure 2 represents a correlation of $3 \div \sqrt{(4 \times 9)} = \cdot 5$. No attempt is made in the diagrams to make the actual areas proportional to the parts of the variance, it is the numbers written in each cell which matter.

The four abilities represented by four tests can clearly overlap in a complicated way, as in Figure 3, which shows one part of the variance (marked *g*) common to all four of the tests ; four parts (left unshaded) each common to three tests ; six parts (shaded) each common to two tests ; and four outer parts (marked *s*) each specific to one test only. The early Theory of Two Factors adopted the hypothesis that, except for very similar varieties of the one test, none of the cells of such a diagram had any contents save those marked *g* and *s*, the general and the specific factors. The "variance" of each ability was in that theory completely accounted for by the variance due to *g*, and the variance due to *s*.

7. *Tetrad-differences.*—In Section 3 it was explained that the discovery made by Professor Spearman was that the correlation coefficients in two columns tend to be in the same ratio as we go up and down the pair of columns. That is to say, if we take the columns belonging to Tests *b* and *f*, and fix our attention on the correlations which *b* and *f* make with *d* and *e*, we have :

	b	*f*
d	·72	·45
e	·56	·35

where

$$\frac{\cdot 72}{\cdot 45} = \frac{\cdot 56}{\cdot 35}$$

This may be written—

$$\cdot 72 \times \cdot 35 - \cdot 45 \times \cdot 56 = 0$$

and in this form is called a "tetrad-difference." In symbols this one is—

$$r_{db}r_{ef} - r_{df}r_{eb} = 0$$

Spearman's discovery may therefore be put thus: "The tetrad-differences are, or tend to be, zero." It is clear that this will be so if, as we said was the case in the Theory of Two Factors, each correlation is the product of two correlations with *g*. For then the above tetrad-difference becomes—

$$r_{dg}r_{bg}r_{eg}r_{fg} - r_{dg}r_{fg}r_{eg}r_{bg}$$

which is identically zero. The present-day test for hierarchical order in a correlation matrix is to calculate all the tetrad-differences (always avoiding the main diagonal) and see if they are sufficiently small. If they are, then the correlations can be explained by a diagram of the same nature as Figure 3, by one general factor and specifics. It is, of course, not to be expected in actual experimenting that the tetrad-differences will be *exactly* zero; no experiment on human material can be as accurate as that. What *is* required is that they shall be clustered round zero in a narrow curve, falling off steadily in frequency as zero is departed from. The number of tetrad-differences increases very rapidly as the number of tests grows, and in an actual experimental battery the tetrads are very numerous indeed. In the small portion of a real correlation table given above (page 7), with six tests, there are 45 tetrad-differences,* and in this instance they are distributed as follows (taking absolute values only and disregarding signs, which can be changed by altering the order of the tests):

From ·0000 to ·0999, 28 tetrad-differences.
From ·1000 to ·1999, 13 tetrad-differences.
From ·2000 to ·2796, 4 tetrad-differences.

This distribution of tetrads can be represented by a "histogram" like that shown in Figure 4, which explains itself. It is clear that some criterion is required by which we can know whether the distribution of tetrad-differences, after they have been calculated, is narrow enough to justify us in assuming the Theory of Two Factors. This criterion

* Not all independent.

is explained in Chapter III, page 41. One form of it consists in drawing a distribution curve to which, on grounds of sampling, the tetrad-differences may be expected to conform. Any tetrad-differences which seem to be too large to be accounted for by the Theory of Two Factors are then examined, to see whether the tests giving them have any special points of resemblance, in content, method, or otherwise, which may explain why they disturb the hierarchy.

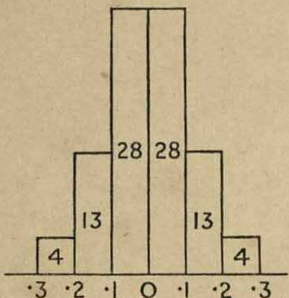


Figure 4.

8. *Group factors.*—As time went on it became clear that the tendency to zero tetrad-differences, though strong, was not universal enough to permit an explanation of *all* correlations between tests in terms of *g* and specifics, with a few slight "disturbers" in the form of slightly overlapping specifics. It became necessary to call in *group factors*, which run through many though not through all tests, to explain the deviations from strict hierarchical order. The Spearman school of experimenters, however, tend always to explain as much as possible by one central factor, and to use group factors only when necessitated. They take the point of view that a group factor must, as it were, establish its right to existence, that the onus of proof is on him who asserts a group factor. As a tiny artificial illustration, a matrix of correlation coefficients :

	1	2	3	4
1	.	·5	·5	·5
2	·5	.	·8	·5
3	·5	·8	.	·5
4	·5	·5	·5	.

would be examined, and its three tetrad-differences found to be :

zero
·15
·15

Inspection shows that the correlation r_{23} is the cause of the discrepancies from zero, and the experimenter trained in the Two-factor school would therefore explain these correlations by a central factor running through them all, plus a special link joining Tests 2 and 3, as in Figure 5.

There are innumerable other possible ways of explaining these same correlations. For example, the linkages between the tests might be as in Figure 6, which gives exactly the same correlations. This lack of uniqueness is something which must always be borne in mind in studying factorial analysis. There are always, as here, innumerable possible analyses, and the final decision between them has to be made on some other grounds. The decision may be psychological, as when for example in the above case an experimenter chooses one of the possible diagrams because it best agrees with his psychological ideas about the tests. Or the decision may be made on the ground that we should be parsimonious in our invention of "factors," and that where one general and one group factor will serve we should not invent five group factors as required by Figure 6. Both diagrams, however, fit the correlational facts exactly, and so also would hundreds of other diagrams which might be made. As has been said, the two-factor tendency is to take the diagram with the largest general factor (and the largest specifics also) and with as few group factors as possible.

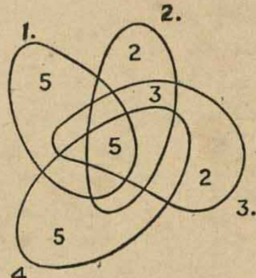


Figure 5.

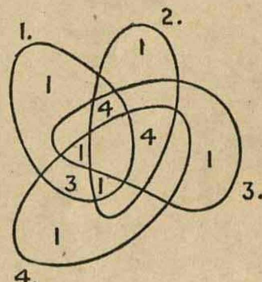


Figure 6.

9. *The verbal factor.*—In this way the Theory of Two Factors has gradually extended the "two" to include, in addition to *g* and specifics, a number of other group factors, still, however, comparatively few. These group factors

bear such names as the *verbal* factor v, a *mechanical* factor m, an *arithmetic* factor, *perseveration*, etc. The characteristic method of the Two-factor school can be well seen, without any technical difficulties unduly obscuring the situation, in the search for a verbal factor. The idea that, in addition to a man's g (which is generally thought of as something innate) there may be an acquired factor of verbal facility which enables him to do well in certain tests, is a not unnatural one. A battery of tests can be assembled of which half do, and half do not, employ words in their construction or solution. The correlation matrix will then have four quadrants, the quadrant V containing the correlations of the verbal tests among themselves, the

V	C
C	P

quadrant P the correlations of the non-verbal or, say, pictorial tests, and the quadrants C containing the cross-correlations of the one kind of test with the other. If the whole table is sufficiently "hierarchical," there is no evidence for a group factor v or a group factor p. If either of these factors exists, there will be differences to be noticed between the six kinds of tetrad which can be chosen, namely :

(1)

	p	p
v	.	.
v	.	.

(2)

	v	v
v	x	x
v	x	x

(3)

	p	p
p	x	x
p	x	x

(4)

	v	p
v	x	.
v	x	.

(5)

	v	p
p	.	x
p	.	x

(6)

	v	p
v	x	.
p	.	x

A tetrad like 1, with two verbal tests along one margin and two pictorial tests along the other, will be found in quadrant *C*. Neither a factor common to the verbal tests only, nor one common to the pictorial tests only, will add anything to any of the four correlations in such a tetrad-difference, which may be expected, therefore, to tend to be zero. If the tetrads in *C* seem to do so, the other tetrads can be examined. Tetrad 2 is taken wholly from the *V* quadrant. In it the verbal factor, if any is present, will reinforce all the four correlations, and should not therefore disturb very much the tendency to a zero tetrad-difference. (Reinforced correlations are marked by x in the diagrams.) The same is true of Tetrad 3 taken wholly from the *P* quadrant. Tetrads 4 and 5 have each two of their correlations reinforced, by the v factor in 4 and by the p factor in 5, but in each case in such a way as not to change very much the tetrad-difference. It is when we come to tetrads like 6, which have one correlation in each of the four quadrants, that the presence of either or both factors should show itself strongly : for the two reinforced correlations here occur on a diagonal, and inflate only the one member of the tetrad-difference—

$$r_{vv}r_{pp} - r_{vp}r_{pv}$$

If, then, a verbal factor, and also a pictorial factor, are present, the tendency for the tetrad-differences to vanish should become less and less strong as we consider tetrads of the kinds 1, 2 and 3, 4 and 5, and especially 6, where the tetrad-differences should leap up. If only the verbal factor is present, tetrad-differences of the kind 3 should vanish rather more than those of the kind 2. But it will not be easy to distinguish between either suspected factor, and both. Tetrads like 6, however, should give conclusive evidence of the presence of one or the other, if not both. Methods like this were employed by Miss Davey (Davey, 1926), who found a group factor, but not one running through *all* the verbal tests, and by Dr. Stephenson (Stephenson, 1931), whose results indicated the presence of a verbal factor.*

* T. L. Kelley had already found by other methods strong evidence of a verbal factor (Kelley, 1928, 104, 121 *et passim*).

10. *Group-factor saturations.*—Just as the g saturations of tests can be calculated, so also can the saturation of a test with any group factor it may contain. The general method of the Two-factor school is first to work with batteries of tests which give no unduly large tetrad-differences, and which also appear to satisfy one's general impression that they test intelligence. From such a battery, of which the best example is that of Brown and Stephenson (B. and S., 1933), the g saturations can be calculated.* Each test has, however, also its specific, which, *so long as it is in the hierarchical battery,* is unique to it and shared with no other member of the battery. A test may now be associated with some other battery of different tests, and with some of these it may share a part of its former specific, as a group factor which will increase its correlation beyond that caused by g. The excess correlation enables the saturation of the test with this group factor to be found—the details are too technical for this chapter—and the specific saturation correspondingly reduced. Finally, the tester may be able to give the composition of a test as, let us say (to invent an example)—

$$\cdot 71g + \cdot 40v + \cdot 34n + \cdot 47s$$

where g is Spearman's g, v is Stephenson's verbal factor, n is a number factor, and s is the remaining specific of the test. The coefficients are the "saturations" of the test with each of these; that is, the correlations believed to exist between the test and these fictitious tests called factors. The squares of these saturations represent the fractions of the test-variance contributed by each factor, and these squares sum to unity, thus:

	Saturation Squared
g . . .	·5041
v . . .	·1600
n . . .	·1156
s . . .	·2209
	1·0006

* For the sake of clarity the text here rather oversimplifies the situation. The battery of Brown and Stephenson contains in fact a rather large group factor as well as g and specifics.

CHAPTER II

BIFACTOR ANALYSIS AND CLUSTERS

1. *The bifactor method.*—Holzinger's Bifactor Method (Holzinger, 1935, 1937*a*) may be looked upon as another natural extension of the simple Two-factor plan of analysis. It endeavours to analyse a battery of tests into one general factor and a number of *mutually exclusive* group factors. A diagram of such an analysis looks like a " hollow staircase," thus :

Test	*g*	*h*	*k*	*l*
1	×	×		
2	×	×		
3	×	×		
4	×		×	
5	×		×	
6	×		×	
7	×			×
8	×			×
9	×			×

Here factor *g* runs through all, as is indicated by the column of crosses. Factors *h*, *k*, and *l* run through mutually exclusive groups of tests each. The saturations with *g* can be calculated from sub-batteries of tests which form perfect hierarchies, by selecting only one test from each group (in every possible way). After these are known, the correlation due to *g* can be removed, and then the saturations due to each group factor found.

The following artificial example will illustrate some of the points of this method. Consider these correlations, which to save space are printed without their decimal points :

	1	2	3	4	5	6	7	8	9	10	11	12
1		57	40	45	63	63	20	28	74	52	45	34
2	57		34	25	53	39	17	44	68	43	39	56
3	40	34		18	57	27	59	16	44	70	73	20
4	45	25	18		27	51	09	12	32	22	20	15
5	63	53	57	27		42	40	26	68	67	63	31
6	63	39	27	51	42		13	18	50	34	30	23
7	20	17	59	09	40	13		08	22	60	64	10
8	28	44	16	12	26	18	08		35	21	18	43
9	74	68	44	32	68	50	22	35		56	50	44
10	52	43	70	22	67	34	60	21	56		78	25
11	45	39	73	20	63	30	64	18	50	78		23
12	34	56	20	15	31	23	10	43	44	25	23	

There are two stages in a bifactor analysis. The first problem is to decide how to group the tests so that those are brought together which share a second or group factor. Then the best method of calculating is needed to find the loadings.

The grouping can partly be done subjectively by considering the nature of each test and putting together memory tests, or tests involving number, and so on. Holzinger uses a " coefficient of belonging," B, to determine the coherence of a group. B is equal to the average of the intercorrelations of the group divided by their average correlation with the other tests in the battery. The higher B is, the more the group is distinguishable as a group. He begins with a pair of tests which correlate highly with one another, and finds their B. Then he adds a third test and finds the B of the three. Then another and another, until B drops too low. There is no fixed threshold for B, but a rather sudden drop would indicate the end of a group.

2. *Tryon's grouping.*—Another plan is to make a graph or profile of each row of correlations and compare these (Tryon, 1939), grouping together those tests with similar profiles. I find it easier to consider only the peaks of each row and compare the rows with regard to these. If we mark, in each row of the above, the five highest correlations

in that row, and also the diagonal cell, we get the following set of peaks :

	1	2	3	4	5	6	7	8	9	10	11	12
1	×	×			×	×			×	×		
2	×	×			×			×	×			×
3			×		×		×		×	×	×	
4	×	×		×	×	×			×			
5	×		×		×				×	×	×	
6	×	×		×	×	×			×			
7			×		×		×		×	×	×	
8	×	×			×			×	×			×
9	×	×			×	?			×	×	?	
10			×		×		×		×	×	×	
11			×		×		×		×	×	×	
12	×	×			×			×	×			×

We then see that, in the rows,

(*a*) Tests 3, 7, 10, 11 have identical peaks,
(*b*) ,, 2, 8, 12 ,, ,, ,,
(*c*) ,, 4, 6 ,, ,, ,,

and we take these as nuclei for three groups. There remain Tests 1, 5, and 9. Their average correlations with each of the above nuclei are :

	a	*b*	*c*
1	·39	·40	·54
5	·57	·37	·35
9	·43	·49	·41

We therefore add Test 1 to group *c*, Test 5 to group *a*, and (less certainly) Test 9 to group *b*. We then rewrite our matrix with the tests thus grouped (see next page) :

It will be seen that certain additions have been made in readiness for the various methods of calculation of the *g* loadings which are then possible. If we symbolize the table overleaf as

A	D	E
D	B	F
E	F	C

	3	5	7	10	11		2	8	9	12		1	4	6	
3		57	59	70	73		34	16	44	20	1·14	40	18	27	·85
5	57		40	67	63		53	26	68	31	1·78	63	27	42	1·32
7	59	40		60	64		17	08	22	10	·57	20	09	13	·42
10	70	67	60		78		43	21	56	25	1·45	52	22	34	1·08
11	73	63	64	78			39	18	50	23	1·30	45	20	30	·95
											6·24				4·62
2	34	53	17	43	39	1·86		44	68	56		57	25	39	1·21
8	16	26	08	21	18	·89	44		35	43		28	12	18	·58
9	44	68	22	56	50	2·40	68	35		44		74	32	50	1·56
12	20	31	10	25	23	1·09	56	43	44			34	15	23	·72
						6·24									4·07
1	40	63	20	52	45	2·20	57	28	74	34	1·93		45	63	
4	18	27	09	22	20	·96	25	12	32	15	·84	45		51	
6	27	42	13	34	30	1·46	39	18	50	23	1·30	63	51		
						4·62					4·07				

all methods depend on using only the correlations in the rectangles D, E, and F, since the suspected group factors which increase the correlations in A, in B, and in C do not influence D, E, and F. Each correlation in the latter rectangles is therefore the product of two *g*-saturations (see page 9). Thus :

$$r_{31} = \cdot 40 = l_3 l_1$$
$$r_{32} = \cdot 34 = l_3 l_2$$
$$r_{12} = \cdot 57 = l_1 l_2$$
$$\therefore\ l_3^2 = \frac{\cdot 40 \times \cdot 34}{\cdot 57} = \cdot 24,\ l_3 = \cdot 49$$

where it should be noted that the three correlations come from E, D, and F respectively.

But this value for the loading of Test 3 depends upon three correlations only and would, in a real experimental set of data, vary somewhat with our choice of the three. A method of using all the possible correlations in these three rectangles is needed. One such is given by Holzinger in his Manual (1937*a*).

3. *Holzinger's formula.*—If all possible ways of choosing the two other tests are taken, and the fraction $\frac{r_{3i}r_{3j}}{r_{ij}}$ formed in each case; and if the numerators of these fractions are added together to form a global numerator, and their denominators to form a global denominator; it will then be found that the fraction thus formed is equal to

$$l_3^2 = \frac{1{\cdot}14 \times {\cdot}85}{4{\cdot}07} = {\cdot}24,\ l_3 = {\cdot}49$$

and this time all available correlations have been used. The rule is to multiply the two totals in the row of the test (1·14 × ·85) and divide by the grand total of the block formed by the other tests concerned (1, 4, and 6 with 2, 8, 9, and 12, i.e. 4·07). For Test 2 this rule gives

$$l_2^2 = \frac{1{\cdot}86 \times 1{\cdot}21}{4{\cdot}62} = {\cdot}49,\ l_2 = {\cdot}70.$$

This Holzinger method is not difficult to extend to four or more groups. If we symbolize a four-group matrix by

A	D	E	G
D	B	F	H
E	F	C	K
G	H	K	L

and consider the first test, then its g-loading l is given by

$$l^2 = \frac{de + dg + eg}{\mathrm{F} + \mathrm{H} + \mathrm{K}}$$

where d, e, and g are the sums of its row in D, E, and G.

4. *Burt's formula.*—Another method is given by Burt (1940, 478). For the numerator of each g loading he takes the *sum* of the side totals which Holzinger multiplied. Thus the numerators are:

for Test 3, 1·14 + ·85 = 1·99
" " 5, 1·78 + 1·32 = 3·10
. . . .
" " 2, 1·86 + 1·21 = 3·07
" " 12, 1·09 + ·72 = 1·81
. . . .
" " 6, 1·46 + 1·30 = 2·76.

The denominators differ in group *a*, group *b*, and group *c*, but all are formed from the three quantities 6·24, 4·62, and 4·07. For group *a* the denominator is :

$$\sqrt{4{\cdot}07}\left\{\sqrt{\frac{6{\cdot}24}{4{\cdot}62}}+\sqrt{\frac{4{\cdot}62}{6{\cdot}24}}\right\}=4{\cdot}08.$$

It will be seen that the two quantities within the curly brackets are the totals of D and E, the two rectangles from which the numerators of group *a* come. By analogy the reader can write down the denominators of group *b* and group *c*—they come to 4·40 and 5·01. Dividing the numerators by the appropriate denominators, we get for the *g* loadings :

Test	3	5	7	10	11	2	8	9	12	1	4	6
g Loading	·49	·76	·24	·62	·55	·70	·33	·90	·41	·82	·36	·55

The proof of Burt's formula is surprisingly easy. If the reader will write down, in place of the correlations in D, E, and F, the literal symbols $l_i l_k$ (for r_{ik})—since our hypothesis is that only *g* is concerned in these correlations—and will write out the sums, etc., of the above calculation literally, he will find that Burt's formula simplifies almost immediately to one *l*, that of the test in question. Burt only gives his formula for three groups. It can be extended to the case of more groups, but becomes cumbersome and rather unwieldy.

5. *The test of correct grouping.*—Now comes the test of whether our grouping is correct, and our hypothesis valid that groups *a*, *b*, and *c* have nothing in common but the factor *g*. Using the loadings we have found, form all the products $l_i l_k$ and subtract them from the experimental correlations. All the correlations in D, E, and F should then vanish or, in a real set of data (ours are artificial), become insignificant. There should, however, remain residues in A, B, and C due to the second factors running through groups *a*, *b*, and *c* respectively. In our example the subtraction of the quantities $l_i l_k$ gives the residues shown at the top of page 25.

The correlations left in A, if they are due to only one other factor (now that *g* has been removed), ought to show

		3	5	7	10	11	2	8	9	12	1	4	6
g Loadings		49	76	24	62	55	70	33	90	41	82	36	55
3	49		20	47	40	46							
5	76	20		22	20	21							
7	24	47	22		45	51							
10	62	40	20	45		44							
11	55	46	21	51	44								
2	70							23	63	29			
8	33						23		30	14			
9	90						63	30		37			
12	41						29	14	37				
1	82											15	18
4	36										15		31
6	55										18	31	

zero or very small tetrads ; and so they do. Those in B are also hierarchical. Those in C are too few to form a tetrad. The second factor in each of these submatrices can now be found in the same way as *g* is found from a matrix with no other factor : see page 9 and, later in this book, pages 42 to 44. The reader should complete the calculation, and will find these loadings :

Factors

Test	*g*	*u*	*v*	*w*
3	·49	·65	.	.
5	·76	·30	.	.
7	·24	·72	.	.
10	·62	·62	.	.
11	·55	·71	.	.
2	·70	.	·44	.
8	·33	.	·47	.
9	·90	.	·11	.
12	·41	.	·62	.
1	·82	.	.	·29
4	·36	.	.	·50
6	·55	.	.	·62

An actual set of data will not give so perfect a hollow staircase, but at this stage the strict bifactor hypothesis can be departed from and additional small loadings or further factors added to perfect the analysis. Where a bifactor pattern exists, a simple method of extracting correlated or oblique factors has been given by Holzinger (1944) "based on the idea that the centroid pattern coefficients for the sections of approximately unit rank may be interpreted as structure values for the entire matrix."

6. *Cluster analysis.*—This is connected with the bifactor method, which is possible when clusters do not overlap. But it is by no means rare to find two or three variables entering into several distinct clusters. Raymond Cattell's article (1944*a*) describes four methods of determining clusters, and gives references which will lead the interested reader back to much of the previous work, and see also Tryon's work *Cluster Analysis*, 1939. The most naïve method of classifying tests into clusters, one needing no mathematics whatever, is simply to put together all the tests which intercorrelate above a certain level. We can illustrate this adequately on the above example. Let us collect into clusters tests which correlate with one another at least 0·40. A routine is desirable to ease the task and avoid overlooking any clusters. Turn to the table on page 20 and write down from the first row all the tests which have correlations of 0·40 or more with Test 1, including itself.

1	2	3	4	5	6	9	10	11
	2			5		9	10	
				5		9	10	
						9	10	

Cluster A, Tests 1, 2, 5, 9, 10.

Then consider the test next to No. 1 in this line, which happens to be Test 2, and go along its line in the correlation table to see which of the tests already noted also correlates sufficiently with Test 2. They are 5, 9, and 10. The other tests of our first line drop out. We then look along the line of Test 5's correlation coefficients, and find that Tests 9 and 10 survive this scrutiny. Finally, we note that Tests 9 and 10 themselves correlate enough. The cluster A is therefore (reading down the left-hand edge of the above triangular set of notes) composed of Tests 1, 2, 5, 9, and 10. At this point, to avoid missing other clusters which may begin with Test 1, it is necessary to consider what would have happened had Test 2 not been in the battery. It would be tedious to describe the whole procedure here, but the reader is urged to go through it, when he will find six clusters, shown in this diagram.

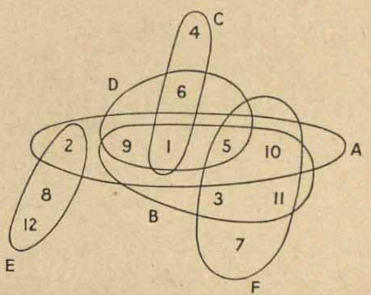


Figure 7.

7. *Comparison with the bifactor groups.*—If we compare these clusters with the grouping we found by Tryon's method of profiles (or peaks), we see that our present clusters F, E, and C are those we arrived at formerly (except for the absence of Test 9 from cluster E). And we notice also that in our diagram these are mutually exclusive clusters. The missing Test 9 is the one we formerly had most doubt about classifying. The reason can be seen from the analysis we have already made. It is highly saturated with the general factor, and only very weakly with the verbal factor which decides its bifactor group.

8. *A less artificial example.*—The above example was an artificial one, made so as to "come out" exactly. Let us turn to a more realistic example where this is not the case. The following correlations—decimal points are again omitted—are from an actual report, but to obviate some embarrassments in a didactic example I have made all the

coefficients rather larger than they actually were. The first seven "tests" are examinations in school subjects, the next four are "non-verbal" tests with simple pieces of apparatus, and the last three are special tests supposed to be uncontaminated by any group factor other than *g*, *v*, and *k* (the "space" factor).

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1 Physics .		76	82	68	64	40	28	44	19	16	21	45	11	10
2 Chemistry .	76		68	62	52	26	26	43	36	29	23	38	15	13
3 Mathematics	82	68		68	47	48	21	37	23	13	20	43	19	18
4 French .	68	62	68		45	23	34	29	25	−13	05	26	34	00
5 Mech. Draw.	64	52	47	45		36	17	53	55	38	21	36	07	42
6 Problems .	40	26	48	23	36		19	51	47	20	40	47	05	36
7 Reading .	28	26	21	34	17	19		09	07	02	17	−07	38	03
8 Koh's Blocks	44	43	37	29	53	51	09		81	50	50	64	43	65
9 Cube Constr.	19	36	23	25	55	47	07	81		42	53	53	37	66
10 Form Board	16	29	13	−13	38	20	02	50	42		52	34	19	38
11 Passalong .	21	23	20	05	21	40	17	50	53	52		32	32	46
12 *g* test .	45	38	43	26	36	47	−07	64	53	34	32		40	57
13 *v* test .	11	15	19	34	07	05	38	43	37	19	32	40		45
14 *k* test .	10	13	18	00	42	36	03	65	66	38	46	57	45	

When by the above method we sort these tests into clusters, using 0·40 as boundary line, we obtain the following diagram :

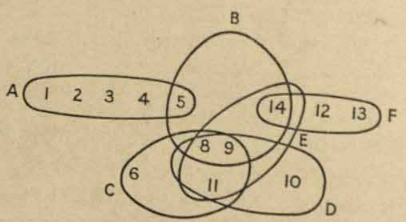


Figure 8.

In passing, we may note that this diagram illustrates what Raymond Cattell (1946) calls a *nuclear* cluster, i.e. one which forms the centre of a number of larger looser clusters. Here the pair 8 and 9 are never separated, occur together in clusters B, C, D, and E, and are such a nuclear cluster. For bifactor analysis, however, we want *non*-overlapping clusters.

9. *A first attempt at grouping.*—Searching in this diagram for at least three non-overlapping contours, we find clusters A, F, and either C or D. Of the alternatives let us take D, and rewrite our table of correlations with these clusters separated. This leaves Tests 6 and 7 out of the picture, and further study of the diagram leads us also to omit 5, which is linked with both F and D through cluster B. Our table, and its calculations, then is as follows :

	1	2	3	4		8	9	10	11		12	13	14	
1		76	82	68		44	19	16	21	1·00	45	11	10	·66
2	76		68	62		43	36	29	23	1·31	38	15	13	·66
3	82	68		68		37	23	13	20	·93	43	19	18	·80
4	68	62	68			29	25	−13	05	·46	26	34	00	·60
										3·70				2·72
8	44	43	37	29	1·53		81	50	50		64	43	65	1·72
9	19	36	23	25	1·03	81		42	53		53	37	66	1·56
10	16	29	13	−13	·45	50	42		52		34	19	38	·91
11	21	23	20	05	·69	50	53	52			32	32	46	1·10
					3·70									5·29
12	45	38	43	26	1·52	64	53	34	32	1·83		40	57	
13	11	15	19	34	·79	43	37	19	32	1·31	40		45	
14	10	13	18	00	·41	65	66	38	46	2·15	57	45		
					2·72					5·29				

From this table, by Holzinger's formula, we obtain the *g* loadings shown at the right of the next table. For example :

$$l_{10}^2 = \frac{0{\cdot}45 \times 0{\cdot}91}{2{\cdot}72} = {\cdot}15055,\ l_{10} = {\cdot}388$$

When, using these *g* loadings, we remove the parts of the correlations due to that factor, we get the following table of residues. For example :

$${\cdot}76 - {\cdot}353 \times {\cdot}404 = {\cdot}62.$$

Residues

	1	2	3	4	8	9	10	11	12	13	14	*g Loadings*
1		62	69	60	09	−08	02	02	14	−08	−07	·353
2	62		53	53	03	05	13	02	03	−06	−07	·404
3	69	53		59	00	−06	−02	00	10	−01	00	·375
4	60	53	59		07	07	−22	−07	06	22	−11	·228
8	09	03	00	07		05	12	−02	−21	−09	17	·984
9	−08	05	−06	07	05		12	12	−14	−04	28	·769
10	02	13	−02	−22	12	12		32	00	−02	19	·388
11	02	02	00	−07	−02	12	32		−14	04	20	·528
12	14	03	10	06	−21	−14	00	−14		−06	15	·867
13	−08	−06	−01	22	−09	−04	−02	04	−06		19	·529
14	−07	−07	00	−11	17	28	19	20	15	19		·488

On examining these residues, however, we see that this time our hypothesis, that the clusters are exclusive with regard to their second group factors, is not justified. True, many of the residues in the side squares are very small. But two facts strike the eye : Test 14 (the *k* or space factor test) has quite large residues with the middle or non-verbal group, and Tests 10 and 11 (Form Board and Passalong) have a much larger residue than the other tests in the middle square. These facts suggest further purging the battery of 14 and either 10 or 11. It is very

Residues

	1	2	3	4	8	9	11	12	13	*g Loadings*
1		57	64	52	08	−08	02	10	−11	·424
2	57		48	45	05	07	03	00	−09	·455
3	64	48		52	00	−05	01	07	−04	·436
4	52	45	52		−02	02	−11	−05	15	·368
8	08	05	00	−02		28	13	−06	−01	·842
9	−08	07	−05	02	28		25	00	04	·633
11	02	03	01	−11	13	25		−04	09	·437
12	10	00	07	−05	−06	00	−04		13	·835
13	−11	−09	−04	15	−01	04	09	13		·522

frequently necessary to "purge" a battery before the proper loadings of the remaining tests can be ascertained.

10. *The purged battery.*—When we do this (the reader should rewrite the tables and carry out the work), we get the loadings and residues shown at the foot of page 30.

This table is much more like our artificial model. None of the correlation coefficients in the side squares are far from zero—we shall learn later how to decide whether they are, in fact, small enough to be ignored. Meanwhile, let us assume this, and suppose, that is to say, that these three groups of tests really are exclusive of one another in their second group factors. Their loadings in these we could then proceed to calculate. This is easily done in the middle group, where there are exactly three tests. We have :

$$m_8^2 = \frac{\cdot 28 \times \cdot 13}{\cdot 25} = \cdot 1456, \; m_8 = \cdot 382$$

$$m_9^2 = \frac{\cdot 28 \times \cdot 25}{\cdot 13} = \cdot 5384, \; m_9 = \cdot 734$$

$$m_{11}^2 = \frac{\cdot 25 \times \cdot 13}{\cdot 28} = \cdot 1161, \; m_{11} = \cdot 341$$

The equations of these three tests are therefore:

$$\begin{aligned} z_8 &= \cdot 842g + \cdot 382h + \cdot 383\, s_8 \\ z_9 &= \cdot 633g + \cdot 734 \;\; + \cdot 246\, s_9 \\ z_{11} &= \cdot 437g + \cdot 341h + \cdot 832\, s_{11} \end{aligned}$$

where the group factor common to them is given the non-committal name h. The coefficients of the specifics are settled by the fact that the sum of the squares of the coefficients of such an equation (since the factors are independent) must equal unity. It will be noticed that Test 11 (Passalong) has here a large specific. It probably shares a good deal of this with Test 10 (Form Board) which we excluded from the battery meanwhile for this very reason.* We cannot similarly calculate the group factor loadings of the third group of tests, for there are only two of them and

* It should be repeated at this point that this example is purely illustrative, and no conclusions about actual tests may be drawn from this or from any of our examples. This is a book about factorial methods, not results.

three tests are necessary. We only know that the product of their two group factor loadings is ·13. This emphasizes the necessity, in planning a bifactor battery, to have a sufficient number of tests. There must be at least three groups, and at least three tests in each group.

The first group has four tests, and our first step should be to see whether its tetrad-differences are zero. If they were exactly zero, it would be immaterial which three of the four tests we chose to calculate loadings from. Here the tetrad-differences, though small (·0084, ·0384, ·0468), are not exactly zero. We shall defer to the next chapter (page 43) the question of how to make the best estimate of the loadings under these circumstances, but the reader might care to calculate them from every possible three of the four tests and average the results. Our illustration has served its purpose of bringing to light difficulties which do not exist in an artificial example made to avoid raising them.

CHAPTER III

SAMPLING ERROR AND THE THEORY OF TWO FACTORS

1. *Sampling error*.—The general idea underlying the notion of a sampling error is not a difficult one. Take, for example, the average height of all living Englishmen who are of full age. This could, if need be, be ascertained by the process of measuring every living Englishman of full age. Actually this has never been done, and when anyone makes a statement such as " The average height of Englishmen is 67½ inches," he is basing it upon a sample only. This sample may not be an unbiased one. Indeed, samples of Englishmen whose height has been officially recorded are heavily loaded with certain classes of Englishmen—for example, prisoners in gaol, and unemployed young men joining the army of preconscription days. The average height of such men may well differ from that of all Englishmen. But when we speak of sampling error, we do not mean error due to the sample being *known* to be a biased one. Even if the sample of Englishmen used to find the average height of their race were, as far as could be seen, a perfectly fair sample, containing the proper proportion of all classes of the community and of all adult ages, etc., it yet would not necessarily yield an average exactly equal to that of *all* Englishmen. Several apparent replicas of the sample would yield different averages. It is these differences, between statistics gathered from different but equally good samples, that we mean by sampling errors.

It is worth while calling attention at this point to a general fact which will be found of importance at a later stage of this book. The *true* average height of Englishmen is only so by definition, and does not in principle differ from the average of a sample. We had to define the population we had in mind as " all living Englishmen of full age." This is a perfectly well-marked body of men. But it is itself in its turn only a sample: a sample of all living Europeans, or all living men. It is, indeed, altering daily and hourly as men die or reach the age of 21, and each generation is a sample of those that have been and may be. Those who reach the age of 21 are only some, and therefore only a sample, of those born. And even those born are only a sample of those who might have been born had times been better or had there been no war, or a tax on bachelors. So the idea of sampling is a relative one, and the "complete population" from which we take samples is a matter of definition only. The mathematical problem in connexion with sampling which it is desirable to solve if possible for each statistic is to find the complete law of its distribution when it is derived from each of a large number of samples of a given size. Mathematically this is often very difficult, and frequently we have to be content with a formula which gives its approximate variance if certain assumptions are allowed and certain small quantities are neglected.

Sampling problems are of two kinds, direct and inverse. The easier kind of problem is to say what the distribution of a statistic will be in samples of a given size when we know all about the true values in the whole population: the more difficult kind is to estimate what the true value of a statistic is in a complete population when we know its observed value in certain samples. They differ as do problems of interpolation and extrapolation. As an example of the direct kind of problem, let us suppose that we actually knew the height of every adult Englishman of full age. We could then, on being told a certain sample of p Englishmen averaged such and such a height, calculate the probability that this sample was a random sample, a probability that would obviously grow less as the average of the sample departed from the average of the whole population. It would also depend on the size of the sample, for if a very large sample deviates far from the true average, it is less likely to be random, more likely to have some reason for the difference, than a small sample with the same average would have.

2. *Standard errors.*—By the distribution of a certain

variable in the population we mean the curve (usually expressed as an equation) showing its frequency of occurrence for each possible value. Thus the curve in Figure 9 might show the distribution of height in living adult Englishmen, by its height above the base line at each point. More men (represented by the line *MN*) have the *average* height, 67½ inches, than have the height 73 inches, the frequency of the latter being shown by the line *PQ*. The shaded area represents all men whose height is 73 inches or more, and its ratio to the area under the whole curve is the probability that an Englishman taken absolutely at random will have a height of 73 inches or more.

Very often distributions are, at any rate approximately, of a certain shape called the "normal curve." The normal curve has a known equation, it is symmetrical about its mid point, and with the aid of published tables can be drawn accurately (or reproduced arithmetically) if we know the mid point *M* (which is the average of the measurements) and a certain distance *ST* or *S′T* (which is equal to the standard deviation of the measurements).

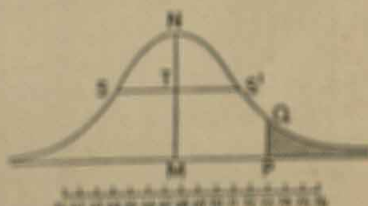


Figure 9

S and *S′* are the points where the curve changes from being convex to being concave.

If the distribution of a variable, say the heights of adult Englishmen, is "normal," then the distribution of the means of samples of *p* Englishmen's heights will also be normal, but will be more closely concentrated about the point *M* than are the measurements of individuals: in point of fact, its variance will be *p* times smaller, its standard deviation thus $\sqrt{p}$ times smaller. That is to say, if we take sample after sample of 25 Englishmen each time, and for each sample record the average height, the means thus accumulated will be distributed in a curve of the same shape as that of Figure 9, but narrower from side to side, so that *SS′* would be one-fifth ($\sqrt{25}$) of what it is in Figure 9, which is the distribution of single measurements.

If a sample were made with some special end in view, such as ascertaining whether red-headed men tend to be tall, we would decide whether we had detected such a tendency by calculating the probability that a mean such as our red-headed sample showed, or a mean still farther away from M, would occur at random. For this purpose we would compare the deviation of our sample from M with the standard deviation of the distribution of such samples, obtained by dividing the standard deviation of individuals by the square root of p, the number in the sample. The ratio of the deviation found, to the standard deviation, is the criterion, and the larger it is the more likely is it that red-headed men really do tend to be tall. For many practical purposes we take a deviation of over twice the standard deviation as "significant."

Sometimes the reader will find significance questions discussed in terms of the "probable error" instead of the standard deviation. The probable error is best considered as a conventional reduction of the standard deviation (or standard error, as it is sometimes called) to two-thirds of its value (more exactly, to ·67449 of its value).

Not only would the average height, or the average weight, of the sample of red-headed men differ from sample to sample. Statistics calculated in more complex ways from the measurements will also vary from sample to sample, as, for example, the variance of height, or the variance of weight, or the correlation of height and weight. Let us consider first the variance of the heights. In the whole population this is calculated by finding the mean, expressing every height as a plus or minus deviation from the mean, squaring all these deviations, and dividing the sum by the number in the population.

This is also how we would find the variance of the sample if we really want the variance *of the sample*. But if we want an estimate of the variance in the whole population, and the sample is small, it is better to divide by one less than the number in the sample. A glimpse of the reason for this can be got by considering the case of the smallest possible sample, namely, one man. Here the mean of the sample is the one height that we have measured, and the deviation of that measurement from the mean of the sample is zero. The formula if we divide by the number in the sample (one) will give zero for the variance—and that is correct for the sample. But it would be too bold to estimate the variance of the whole population from one measurement: if we divide by one less than the sample we get variance $= 0/0$, that is, we don't know, which is a wiser statement.*

More generally we can begin to understand the reason for dividing by $(p - 1)$ instead of by p by the following considerations.

The quantity we want to estimate is the mean square deviation of the measurements of the whole population, the deviations being taken from the mean of that whole population. We do not, however, know that *true mean*, and therefore in a sample we are reduced to using the *mean of the sample*, which except by a miracle will not exactly coincide with the true or population mean. The consequence is that the sum of the squares we obtain is smaller than it would have been had we known and used the true mean. For it is a property of a mean that the sum of the squares of deviations from it is smaller than of deviations from any other point.

* It is important to remember that sampling the population is not the only source of error in the measurement of statistics, e.g. the correlation coefficient. All sorts of influences may disturb it. These will *usually* "attenuate" the correlation coefficient, i.e. tend to bring it nearer to zero, as can be seen when we consider that a perfect correlation only can be reduced by error. But they will not always do so, and if the errors in the two trait measurements are themselves correlated, they may even increase the true correlations in a majority of cases. An estimate of the amount of variable error present can be made from the correlation of two measurements of the same trait on the same group, a correlation called the "reliability," which should be perfect if no variable errors are present. Spearman's correction for attenuation (see Brown and Thomson, 1925, 156) is based upon this. Like all estimates, the correction for attenuation is correct, even if the errors are uncorrelated, only on the average and not in each instance, and it should never be used unless it is small. If it is large, the experiments are "unreliable" and should be improved.

Consider for example the numbers 2, 3, and 7. Their mean is 4, and the sum of the squares about 4 is—

$$(-2)^2 + (-1)^2 + 3^2 = 14$$

About *any* other point this sum will be greater than 14. About 5, for example, the sum is—

$$(-3)^2 + (-2)^2 + 2^2 = 17$$

About 2 the sum is—

$$0^2 + 1^2 + 5^2 = 26$$

It follows that the sum of the squares we obtained by using the sample mean was as small as possible, and in the immense majority of cases smaller than the sum about the true mean. It is to compensate for this that we divide by $(p - 1)$ instead of by p.

These elementary considerations do not of course indicate just why this procedure should, in the long run, exactly compensate for using the sample mean. Why not $(p - 2)$, one might say, or $(p - 3)$? It is not possible, in an elementary account like the present, to answer this. Geometrical considerations, however, throw some further light on the problem. The p measurements of the sample may be thought of as existing in a certain space of $(p - 1)$ dimensions. For example, two points define a line (of one dimension), three points define a plane (of two dimensions), and so on. The true mean of the whole population is not likely to be within that space, whereas the mean of the sample *is*. The deviations we have actually squared and summed are therefore in a space of one dimension less than the space containing the true mean. One " degree of freedom " has been lost by the fact that we have forced the lines we are squaring to exist in a space of $(p - 1)$ dimensions instead of permitting them to project into a p-space. Hence the division by $(p - 1)$ instead of p.

This principle goes farther. For each statistic which we calculate from the sample itself and use in our subsequent calculations, we lose a " degree of freedom."

The standard error of a variance v, if the parent population from which the samples are drawn is normally distributed, is estimated as—

$$\frac{v\sqrt{2}}{\sqrt{(p-1)}}$$

where p is the number of persons in the sample. The standard error of a correlation coefficient r is, with the same condition, estimated as—

$$\frac{1-r^2}{\sqrt{(p-1)}}$$

The use of this standard error, however, should be discontinued (unless the sample is large and r small).

Fisher (1925, page 202) has pointed out that the use of the formula for the standard error of a correlation coefficient is valid only when the number in the sample is large and when the true value of the correlation does not approach ± 1. For in small samples the distribution of r is not normal, and even in large samples it is far from normal for high correlations. The distribution of r for samples from a population where the correlation is zero differs markedly from that where the correlation is, say, 0·8. This means that the use of a standard error for testing the significance of correlation coefficients should, except under the above conditions, be discouraged.

To get over the difficulty Fisher transforms r into a new variable z given by—

$$z = \tfrac{1}{2}\{\log_e(1+r) - \log_e(1-r)\}$$
$$= r + \tfrac{1}{3}r^3 + \tfrac{1}{5}r^5 + \cdots$$

It is not, however, necessary to use this formula, as complete tables have been published for converting values of r into the corresponding values of z. As r goes from -1 to $+1$, z goes from $-\infty$ to $+\infty$, and $r = 0$ corresponds to $z = 0$.

The great advantage of using z as a variable instead of r is that the form of the distribution of z depends very little upon the value of the correlation in the population from which samples are drawn. Though not strictly normal, it tends to normality rapidly as the size of the sample is increased, and even for small samples the assumption of normality is adequate for all practical purposes. The standard deviation of z may in all cases be taken to be $1/\sqrt{p-3}$, where p is the number of persons in the sample.

3. *Error of a single tetrad-difference.*—For our discussion of the influence of sampling on the factorial analysis of tests one of the most important quantities to know is the standard error of the tetrad-difference. There has been much debate concerning the proper formula for this. (See Spearman and Holzinger, 1924, 1925, 1929; Pearson and Moul, 1927; Wishart, 1928; Pearson, Jeffery, and Elderton, 1929; Spearman, 1931.) That generally employed is formula (16) in the Appendix to Spearman's *The Abilities of Man*:

Standard error of $r_{13}r_{24} - r_{23}r_{14} =$

$$\frac{2}{\sqrt{N}}[r^2(1 - r_{12} - r_{34} + r^2) + (1 - 2r^2)s^2]^{\frac{1}{2}}$$ [Spearman and Holzinger's formula (16).]

where N is the number of persons in the sample,*

r is the mean of the four correlation coefficients, and
s^2 is their mean squared deviation (variance) from r.

The probable error is ·6745 times the above. A worked example will be found on page xii of Spearman's Appendix, using (which is all one can do) the *observed* values of the r's.

It will be remembered that in Section 7 of Chapter I we stated Spearman's discovery in the form "tetrad-differences tend to be zero." If tetrad-differences in the whole population, however, were all actually zero, they would not remain exactly zero in samples, and it is only samples that are available to us. We are faced, therefore, with a two-fold problem. (*a*) We have to decide, from the size of the tetrad-differences actually found in our sample, whether the sample is compatible with the theory that the tetrad-differences are zero in the whole population. But (*b*) we should also go on to consider whether the sample is equally compatible with the opposed hypothesis that the

* We use p to mean the number of persons in this book, but are retaining N here and in "formula 16A" below to preserve the usual appearance of these well-known and much-used expressions.

tetrad-differences are not zero in the whole population, leaving a verdict of "not proven.") (See Emmett, 1936.)

4. *Distribution of a group of tetrad-differences.*—The actual calculation, for every separate tetrad-difference, of its standard error by Spearman and Holzinger's formula (16) is, however, an almost impossibly laborious task. In a table of correlations formed from n tests there are $n(n-1)/2$ correlation coefficients, and $n(n-1)(n-2)(n-3)/8$ different (though not independent) tetrad-differences. Any one particular correlation-coefficient is concerned in $(n-2)(n-3)$ different tetrad-differences, and any one test in $(n-1)(n-2)(n-3)/2$ different tetrad-differences. Thus with ten tests there are 630 tetrad-differences, and with twenty tests 14,535 tetrad-differences. In the latter case, any one test is concerned in 2,907. Under these circumstances, it is natural to look for a more wholesale method than that of calculating the standard error of each tetrad-difference. The method adopted by Spearman is to form a table of the distribution of the tetrad-differences, and compare this distribution with that of a normal curve centred at zero and with standard deviation given by—

$$\frac{2}{\sqrt{N}}[r^2(1-r)^2 + (1-R)s^2]^{\frac{1}{2}}$$ [Spearman and Holzinger's formula (16A).]

where N = number of persons in the sample,

r = the mean of all the r's in the whole table,

s^2 = their mean squared deviation from r.

$$R = 3r.\frac{n-4}{n-2} - 2r^2.\frac{n-6}{n-2}, \text{ and}$$

n = number of tests.

Numerous examples of the comparison of "histograms" of tetrad-differences with normal curves whose standard deviation is found by (16A) are given in Spearman's *The Abilities of Man*. This method of establishing the hypothesis, that the tetrad-differences are derived by sampling from a population in which they are really zero, is open to the same doubt as was explained in the simpler case of one tetrad-difference. The comparison can prove that the tetrad-differences observed are compatible with that hypothesis. It does not in itself prove that they are compatible with that hypothesis only; and, as Emmett has shown in the article already mentioned, the odds are commonly rather against this.

The usual practice, moreover, is to "purify" the battery of tests until the actual distribution of tetrad-differences agrees with (16A), so that in effect all that is then proved is that a team *can* be arrived at which *can* be described in terms of two factors. This, although a more modest claim than has often been made, and certainly less than is implicitly understood by the average reader, is nevertheless a matter of some importance. Not all teams of tests can be explained by one common factor; but it is not very difficult to find teams which can. There is little doubt in the minds of most workers that a *tendency* towards hierarchical order actually exists among mental tests.

5. *Spearman's saturation formula.*—It will be remembered from Section 4 of Chapter I that the calculation of the g saturation of each test forms an important part of the Spearman process. We saw there that in a hierarchical matrix each correlation is the product of the two g saturations of the tests, for example—

$$r_{34} = r_{3g} \cdot r_{4g}$$

Since this is so, each g saturation can be calculated from the correlations of a test with two others, and their inter-correlation. Thus to find r_{1g} we can take Tests 2 and 3 as reference tests, when we have—

$$\frac{r_{12}r_{13}}{r_{23}} = \frac{r_{1g}r_{2g} \cdot r_{1g}r_{3g}}{r_{2g} \cdot r_{3g}} = r_{1g}{}^2$$

When the matrix is really hierarchical, and there are no sampling errors present, it is immaterial which two tests we associate with Test 1 in order to find its g saturation. We have, in fact, in that case—

$$\frac{r_{12} \cdot r_{13}}{r_{23}} = \frac{r_{14} \cdot r_{15}}{r_{45}} = \frac{r_{12} \cdot r_{15}}{r_{25}} = etc.$$

But even if the correlations, measured in the whole population, were really exactly hierarchical, sampling errors would make these fractions differ somewhat from one another, and we are faced with the problem of deciding which value to accept for the g saturation. The average of all possible fractions like the above would be one very plausible quantity to take but is laborious to compute. Spearman therefore adopts a fraction—

$$\frac{r_{12} \cdot r_{13} + r_{14} \cdot r_{15} + r_{12} \cdot r_{15} + etc.}{r_{23} + r_{45} + r_{25} + etc.} = r_{1g}^2$$

whose numerator is the sum of the numerators, and whose denominator is the sum of the denominators, of the single fractions. This combined fraction he computes in a tabular manner which we will next describe, by the algebraically equivalent formula—

$$r_{1g}^2 = \frac{A_1^2 - A_1'}{T - 2A_1}$$ [Spearman's formula (21), Appendix, *Abilities of Man.*]

The quantities A_1, A_2, etc., are the sums of the rows (or columns) of the matrix of correlations without any entries in the diagonal cells. (The arithmetical example is confined to five tests to economize space):

	1	2	3	4	5	A	A^2
1	.	·50	·34	·33	·24	1·41	1·988
2	·50	.	·56	·32	·15	1·53	2·341
3	·34	·56	.	·13	·35	1·38	1·904
4	·33	·32	·13	.	·29	1·07	1·145
5	·24	·15	·35	·29	.	1·03	1·C61

$T = 6{\cdot}42$

T is the sum of all the A's, and therefore of all the correlations in the table (where each occurs twice). A new table is now written out, with each coefficient squared, and its rows summed to obtain the quantities A':

	1	2	3	4	5	A'
1	.	·250	·116	·109	·058	·533
2	·250	.	·314	·102	·023	·689
3	·116	·314	.	·017	·123	·570
4	·109	·102	·017	.	·084	·312
5	·058	·023	·123	·084	.	·288

The calculation of all the saturations is then best performed in a tabular manner, thus :

	A^2	A'	$A^2 - A'$	$2A$	$T-2A$	$\dfrac{A^2 - A'}{T-2A}$	g Saturation
1	1·988	·533	1·455	2·82	3·60	·4042	·66
2	2·341	·689	1·652	3·06	3·36	·4917	·70
3	1·904	·570	1·334	2·76	3·66	·3645	·60
4	1·145	·312	·833	2·14	4·28	·1946	·44
5	1·061	·288	·773	2·06	4·36	·1773	·42

where the last column is the square root of the preceding. The reader should calculate the six different values of r_{1g} from the original table by the formula $(r_{1j} \cdot r_{1k}/r_{jk})^{\frac{1}{2}}$, for comparison with the value ·66 obtained above. He will find—

·55	·72	·89
	·93	·48
		·52

with an average of ·68.

6. *Residues.*—If the correlations which would arise from these saturations or loadings are calculated, and subtracted from the observed correlations, we obtain the residues which have then to be examined to see if they are small enough to be attributable to sampling error. In the following double table of correlations are set out the observed correlations uppermost, and those calculated from the g saturations below. The difference is the residue, which may be plus or minus :

g *Loadings*	·66	·70	·60	·44	·42
·66	.	·50 ·46	·34 ·40	·33 ·29	·24 ·28
·70	·50 ·46	.	·56 ·42	·32 ·31	·15 ·29
·60	·34 ·40	·56 ·42	.	·13 ·26	·35 ·25
·44	·33 ·29	·32 ·31	·13 ·26	.	·29 ·18
·42	·24 ·28	·15 ·29	·35 ·25	·29 ·18	.

The lower numbers are the products of the two saturations. In this case the residues range from $-\cdot 14$ to $+\cdot 14$ and at first sight appear in many cases to be too large to be neglected in comparison with the original correlations.

To check this impression, consider the correlation ·56 and the value ·42 from which it is supposed to depart only by sampling error, a deviation of ·14. Fisher's z corresponding to $r = \cdot 42$ is ·45, and that corresponding to $r = \cdot 56$ is $z = \cdot 63$, so that the z deviation is ·18. The standard deviation of z for 50 cases is $1 \div \sqrt{47} = \cdot 15$. The deviation is little larger than one standard deviation and cannot therefore be called significant. But as the reader will observe, this conclusion is due more to the large size of the standard error than to the small size of the residue. The residue is here *attributable* to sampling error, because the latter is so large. But because the latter is large it does not follow that the large residue is certainly due to it.

7. *Reference values for detecting specific correlation.*—If after a calculation like that described, one of the residues is found to be too large to be explicable by sampling error, the excess of correlation over that due to g is attributed to "specific correlation," meaning correlation due to a part of their specific factors being not really unique but shared by these two tests. In the case of our numerical example, if the number of subjects tested had been larger, the standard errors of the coefficients would have been smaller, and some of the discrepancies between the experimental values and those calculated from the g saturations would have been too large to be overlooked, but would have had to be attributed to specific correlation. In such a case, the g loadings would, of course, be wrong and would have to be recalculated from the battery after one of the tests concerned in the specific correlation was removed from it. Later, the other test could be replaced in the battery instead of the first, and thus *its* g saturation found. The difference between the experimental correlation of the two, and the product of their g saturations, with a standard error dependent on the size of the sample, would be then attributed to their specific linkage.

If two tests, v and w, are thus suspected of having a specific link as well as that due to g, it is clear that the smallest battery of tests which could be used in the above manner to detect that link would be one of *two* other tests, x and y, say, to make up a tetrad :

	v	x
w	r_{vw}	r_{xw}
y	r_{vy}	r_{xy}

and these two " reference " tests would have to be known to have no specific links with each other or with the two suspected tests. The example which gave rise to Figure 5 (see Chapter I, page 15) illustrates this. Tests 2 and 3 there are, let us suppose, those with a suspected specific link. The tetrad-difference to be examined by means of Spearman's formula (16) is that which has r_{23} as one corner. In such a case, where the two reference tests 1 and 4 are known to have no link except g with one another, or with the other two tests, two of the possible tetrad-differences ought to be larger than three times the standard error given by formula (16), and equal to one another, while the third tetrad-difference should be zero (or sufficiently near to zero, in practice) (Kelley, 1928, 67).

The g saturation of each of the tests under examination for specific correlation can be found by grouping it with the two reference tests. Thus in the case of our Figure 5, we have—

$$r_{2g}^2 = \frac{r_{12} \cdot r_{24}}{r_{14}} = \frac{\cdot 5 \times \cdot 5}{\cdot 5} = \cdot 5$$

$$r_{3g}^2 = \frac{r_{13} \cdot r_{34}}{r_{14}} = \frac{\cdot 5 \times \cdot 5}{\cdot 5} = \cdot 5$$

Therefore the correlation between 2 and 3 which is due to g is—

$$r_{2g} \cdot r_{3g} = \sqrt{\cdot 5} \times \sqrt{\cdot 5} = \cdot 5$$

and the difference between this and ·8, the actual value, is the part to be explained by the specific factor shared by these two tests.

When there are several reference tests available, all believed to have no link except *g* with one another or with the two tests suspected of specific overlap, there will be a number of ways of picking two of them to obtain the tetrad required to decide the matter, and the results will, because of sampling and other errors, be discrepant. Under these circumstances Spearman has devised an interesting procedure for amalgamating the results into one. A numerical example is given by him on page xxii of the Appendix to *The Abilities of Man*.

CHAPTER IV

THE DEFINITION OF *g*

1. *Any three tests define a " g."*—The idea of *g* arose out of Professor Spearman's acute observation that correlation coefficients between tests tend to show hierarchical order: that is, that their tetrad-differences tend to be zero or small; or in more technical terms still, that the rank to which a matrix of correlation coefficients can be " reduced " by suitable diagonal elements tends towards rank *one*. This fundamental fact is at the basis of all those methods of factorial analysis which magnify specific factors. In consequence, *correlation coefficients* between a number of variables can be adequately accounted for by a few common factors. To be adequately described by one only—a *g*—the " reduced " rank of the correlation matrix has to be *one*, within the limits of sampling error.

Suppose now that we have three tests and have, in the whole population, measured their correlation coefficients. If, as is usually the case, these coefficients are all positive, and if each of them is at least as large as the product of the other two, we can explain them by assuming one *g* and three specifics s_1, s_2, and s_3. There are many other ways of explaining them, but let us adopt this one. *We have thereby defined a factor* g *mathematically* (Thomson, 1935*a*, 260). It is then for the psychologist to say, from a consideration of the three tests which define it, what name this factor shall bear and what its psychological description is. The psychologist may think, after studying the tests, that they do not seem to him to have anything in common, or anything worth naming and treating as a factor. That is for him to say. Let us suppose that at any rate he does not reject the possibility, but that he would like an opportunity of studying other tests which (mathematically speaking) contain this factor, and have nothing else in common, before finally deciding.

In that case the experimenter must search for a fourth test which, when added to these three, gives tetrad-differences which are zero ; and then for a fifth and further tests, each of which makes zero tetrad-differences with the tests of the pre-existing battery. This extended battery the experimenter would lay before the psychological judge, to obtain a ruling whether the single common factor, of which it is the now extended but otherwise unaltered definition, is worthy of being named as a psychological factor.

2. *The extended or purified hierarchical battery.*—Mathematically, *any* three tests with which the experimenter cared to begin would define " a " g, if we except temporarily the case, to which we shall later return, of three correlation coefficients, one of which is less than the product of the other two. The experimental tester, however, might in some cases have great difficulty in finding further tests, to add to the original three, which would give zero tetrad-differences. Unless he could do so, it is unlikely that the psychological judge would accept the factor as worthy of a name and separate existence in his thoughts. It is, for example, an experimental fact that starting with three tests which a general consensus of psychological opinion would admit to have only " intelligence " as a common requirement, it has proved possible to extend the battery to comprise about a score of tests without giving any tetrad-differences which cannot be regarded as zero.* Even that has not been accomplished without difficulty, and without certain blemishes in the hierarchy having to be removed by mathematical treatment. But the fact that with these reservations it is possible, and that psychological judgment endorses the opinion that each test of this battery requires " intelligence," is the main evidence behind the actual " existence " of such a factor as " g, general intelli-

* The process of making such a battery of tests to define general intelligence (see Brown and Stephenson, 1933) has not in fact taken the form of choosing three tests as the basal definition and then extending the battery. Instead, a number of tests which, it was thought from previous experience, would act in the desired way have been taken, and the battery thus formed has then been purified by the removal of any tests which broke the hierarchy.

gence." It must be noted that the word " existence " here does not mean that any physical entity exists which can be identified with this *g*. It does mean, however, that, as far as the experimental evidence goes, there is some aspect of the causal background which acts " as if " it were a single unitary factor in these tests.

The important point to note is that the experimenter has produced a battery of tests which is, he claims, hierarchical; that the mathematician assures him that such a battery acts " as if " it had only one factor in common (though it *can* also be explained in many other ways), and that the psychologist agrees that psychologically the existence of such a factor as the sole link in this battery seems a reasonable hypothesis.

3. *Different hierarchies with two tests in common.*—Now, it must be remembered that, starting with three other tests, which may contain two of the former set, it may very well be possible to build up a different hierarchy. Only experiment could show whether this were possible in each case, there is no mathematical difficulty in the way. Such a hierarchy would also define " a " *g*, but this would be usually a different factor from the former *g*. If there were *three* tests common to the two hierarchies, then the two *g*'s could be identified with one another (sampling errors apart), and the three tests would be found to have the same saturations with the one *g* as with the other. But if only two tests were common to the two batteries this would not in general be the case, and the different saturations of these tests with the two *g*'s would show that the latter were different (Thomson, 1935*a*, 261–2). Under such circumstances the psychologist has to choose. He cannot have both these *g*'s. Both are mathematically of equal standing, it is a psychological decision which has to be made. When one *g* is accepted, the other, as a factor, must then be rejected and a more complicated factorial analysis of the second hierarchy has to be built up which is consistent with this.

4. *A test measuring " pure* g."—Although the hierarchical battery defines a *g*, it does not enable it to be measured exactly (but only to be estimated) unless either it contains

an infinite number of tests, or a test can be found which conforms to the hierarchy and has a g saturation of unity.* In the latter case this test which is " pure g " is such that when it is considered along with any other two tests of its hierarchy, its correlations with them, multiplied together, give the intercorrelation of those two with one another: if k is the " pure " test, then—

$$r_{ik}r_{jk} = r_{ij}$$

its g saturation being—

$$\sqrt{\frac{r_{ik}r_{jk}}{r_{ij}}} = 1$$

No such " pure " test of the g which is defined by the Brown-Stephenson hierarchy of nineteen tests has yet been found. Such a pure test, with full g saturation, must not be confused with tests which are sometimes called tests of pure g because they do not contain certain other factors, in particular the verbal factor. Thus the " S.V.P." (Spearman Visual Perception) tests are referred to by Dr. Alexander (1935, 48) as a "pure measure of g"; but their saturations with g are given by him (page 107) as ·757, ·701, and ·736 respectively, so that in each case only about half the variance is " g " and half is a specific.

5. *The Heywood case.*—Consider the case where three tests are such that—

$$r_{ik}r_{jk} > r_{ij}$$

In such a case the g saturation of the test k, if we calculate it, is greater than unity, which is impossible. Yet it is possible, in theory at least, to add tests to such a triplet to form an extended hierarchy with zero tetrad-differences. There can be one such case (but only one) in a hierarchy. We shall call them *Heywood cases*, as this possibility was first pointed out by him (Heywood, 1931). As an artificial example, consider these correlations:

* It is understood, of course, that even such a test would give different measures of a man's g from day to day, if the man's performance in it varied (as it undoubtedly would) from day to day. By measuring with exactness is meant, in this part of the text, measurement free from the uncertainty due to the factors outnumbering the tests. We are assuming sampling errors to be nil.

case of pure g will leave one of the rows of the above sum non-zero. To make the whole sum zero, one case must be a Heywood case, giving—

$$1 - r_{ig}^2 \text{ negative.}$$

It would seem, therefore, that by the time we have added hierarchical tests to make them equal in number to the persons, we will necessarily have added a Heywood hierarchical case (of which there can be only one in a hierarchy). But we have agreed that the discovery of a Heywood case will cause us to abandon the hierarchy as a definition of g!

The case where the number of tests is increased to equal the number of persons may seem to the reader to be an academic case only. But the case of reducing the number of persons until they equal the number of tests is one which could easily be realized in practice, and presents equal theoretical difficulties. This draws attention to the dependence of any definition of factors on the sample of persons tested. If we have a perfect hierarchy of, say, 50 tests, in a population of, say, 1,000 persons, a sample of fifty persons from the above thousand, if it gives hierarchical order, will give a Heywood case, and its g will be impossible.

If the g corresponding to the original analysis on the thousand persons were anything real, such as a given quantity of mental energy available in each person, then it ought always to be possible, one might erroneously think, to find fifty persons and fifty tests to give a hierarchy, without a Heywood case. But that cannot be easily said. It is impossible, from the correlations alone, to distinguish a real g from one imitated by a fortuitous coincidence of specifics. Even if g were a reality, a sample of persons equal in number to the tests could not give a hierarchy without a Heywood case, and their apparent g would be fortuitous.

Now the case of a test of pure g is on the border line of the Heywood cases. It is clear then that it will be suspect, as being probably only fortuitous, if the number of persons does not far exceed the number of tests.

7. *Singly conforming tests.*—There remains one other conceivable method of measuring *g* exactly,* by the use of certain tests which, when they are all present, destroy the hierarchy, although any one of them can enter the battery without marring it—" singly conforming " tests (Thomson, 1934*b* ; and 1935*a*, 253–6). It will be shown in later chapters on factor estimation that the reason factors cannot be measured exactly, but have to be estimated only, is that they outnumber the tests. Every new test which conforms to a hierarchy adds a new specific (unless it is pure *g*), and thus continues the excess of factors over tests. It can occur, however, that the correlation of two tests with each other breaks a hierarchy, although either of them alone conforms otherwise. Such a case occurs in the Brown-Stephenson battery, for example, one of whose correlation coefficients has to be suppressed before the hierarchy is acceptable.

In such a case, if the psychologist is prepared to accept either test as a member of the battery, the erring correlation coefficient must be due to these two tests sharing some portion of their specifics with one another. If, as may happen (apart from error which we are supposing absent), their intercorrelation shows that they have only one specific factor between them, and differ only in their saturations, then they enable the estimate of *g* to be turned into accurate measurement. For example, consider the following matrix of correlations :

	1	2	3	4	5	6
1	.	·669	·592	·458	·335	·251
2	·669	.	·566	·438	·870	·240
3	·592	·566	.	·387	·283	·212
4	·458	·438	·387	.	·219	·164
5	·335	·870	·283	·219	.	·120
6	·251	·240	·212	·164	·120	.

This is a perfect hierarchy except for the correlation—

$$r_{25} = \cdot 870$$

* By " exactly " is meant, with the same exactness as the test scores, without the additional indeterminacy due to an excess of factors over tests.

Every tetrad-difference, which does not contain this correlation, is zero. If either Test 2 or Test 5 is removed from the battery, there remains a perfect hierarchy. If Test 5 is removed, we can calculate from the remaining battery the g saturations :

Test	1	2	3	4	6
g saturation	·837	·800	·707	·548	·300

If we remove Test 2 and restore Test 5, we get the following :

Test	1	3	4	5	6
g saturation	·837	·707	·548	·400	·300

From either hierarchy we can estimate g. The correlation of our estimates with " true g " will be—

$$\sqrt{\frac{S}{S+1}}$$

where

$$S = \Sigma \frac{\text{saturation}^2}{1 - \text{saturation}^2}$$

and we find for the two hierarchies the g correlations of ·92 and ·90.

From the two Tests 2 and 5 alone, however, we can obtain a g correlation of unity.

The reason for this is that the correlation of Tests 2 and 5 is such as to show that their specifics are identical, the two tests differing only in their loadings. Their equations are—

$$z_2 = \cdot 8g + \sqrt{(1 - \cdot 8^2)}s_2$$
$$z_5 = \cdot 4g + \sqrt{(1 - \cdot 4^2)}s_5$$

If the whole of s_2 is identical with the whole of s_5, their intercorrelation should be—

$$\cdot 8 \times \cdot 4 + \sqrt{(1 - \cdot 8^2)(1 - \cdot 4^2)} = \cdot 870$$

and this is its experimental value.

We could, therefore, have seen at the beginning, if we had tested the above fact, that these two tests would make a perfect battery for measuring g. We have the simultaneous equations—

$$z_2 = \cdot 8g + \cdot 6s$$
$$z_5 = \cdot 4g + \cdot 917s$$

from which we can eliminate s.

We see, therefore, that under certain hypothetical circumstances, a more exact estimate of g can be obtained from two of these "singly conforming" tests than the hierarchy with which they conform individually. Those circumstances are, that their correlation with one another (the correlation which breaks the hierarchy because it is too large) should either equal—

$$r_{ig}r_{jg} + \sqrt{(1 - r_{ig}^2)(1 - r_{jg}^2)}$$

or should approach this value.

It cannot in actual practice be expected to equal it, as in our artificial example. For we have disregarded errors, which are sure in some measure to be present. At what stage will the pair of singly conforming tests cease to be a better measure of g than the better of the two hierarchies made by deleting either the one or the other? If in our example the correlation ·870 of Tests 2 and 5 be imagined to sink little by little, the correlation of their estimate with g will sink from unity. The better of the two hierarchies gives a multiple correlation of ·922. When the correlation r_{25} has sunk from ·870 to ·847, these two singly conforming tests will give the same multiple correlation, ·922. If this defect from the full ·870 is due entirely to error, then a fall to ·847 corresponds to reliabilities of the two tests of the order of magnitude of ·98, if they are equally reliable. This is a very high reliability, seldom attained, so that in a case like our example quite a small admixture of error would make the singly conforming tests no better at estimating g than the hierarchy. We are here, however, neglecting the fact that error would also diminish the efficiency of the hierarchy. Nevertheless, the chance of finding a pair of singly conforming tests, highly reliable, and having no specifics except that which they share, seems small, as small as the chance of finding a test of pure g, perhaps. It might possibly turn out, however, that a matrix of several (say t) singly conforming tests would be practicable. Such a set would measure g exactly if among them they added only $t - 1$ new specifics to the hierarchy. Their saturations would be found by placing them one at a time in the hierarchy, and then their regression on g calculated by Aitken's method (see Chapter XIV). The necessity for the hierarchy in the background, in all this, is clear : it is there to assure us that each singly conforming test is compatible with the definition of g, and to enable its g saturation to be calculated.

8. *The danger of " reifying " factors.*—The orthodox view of psychologists trained in the Spearman school is that g is, of all the factors of the mind, the most ubiquitous. " All abilities involve more or less g," Spearman said, although in some the other factors are " so preponderant that, for most purposes, the g factor can be neglected." With this view, the present author has always agreed, provided that g is interpreted as a mathematical entity only, and judgment is suspended as to whether it is anything more than that.

The suggestion, however, that g is " mental energy," of which there is only a limited amount available, but available in any direction, and that the other factors are the neural machines, is one to be considered with caution. The word *energy* has a definite physical meaning. " Mental energy " may convey the meaning that the energy spoken of is the same as physical energy, though devoted to mental uses. If that meaning is accepted, innumerable difficulties follow, not the least being the insoluble questions of the connexion of body and mind, and of freewill versus determinism. A less obscure difficulty is that there seems to be no easily conceivable way in which the " energy " of the whole brain can be used in any direction indifferently, except by the " neural engines " also all taking part. The energy of a neurone seems to reside in it, and the passage of a nerve impulse along a neurone seems to resemble rather the burning of a very rapid fuse, than the conduction of electricity, say, by a wire.

If " mental energy " does not mean physical energy at all, but is only a term coined by analogy to indicate that the mental phenomena take place " as if " there were such a thing as mental energy, these objections largely disappear. Even in physical or biological science, the things which are discussed and which appear to have a very real existence to the scientist, such as " energy," " electron," " neutron," " gene," are recognized by the really capable experimenter as being only manners of speech, easy ways of putting into comparatively concrete terms what are really very abstract ideas. With the bulk of those studying science there exists always the danger that this may be taken too literally, but this danger does not justify us in ceasing to use such terms. In the same way, if terms like " mental energy " prove to be useful, and can be kept in their proper place, they may be justified by their utility. The danger of " reifying " such terms, or such factors as *g*, *v*, etc., is, however, very great.

PART II

MULTIPLE-FACTOR ANALYSIS

CHAPTER V

THE CENTROID METHOD

1. *Need of group factors.*—The two-factor method of analysis, described in an earlier chapter, began with the idea that a matrix of correlations would ordinarily show perfect hierarchical order if care was taken to avoid tests which were "unduly similar," i.e. very similar indeed to one another. If such were found coexisting in the team of tests, the team had to be "purified" by the rejection of one or other of the two. Later it became clear that this process involves the experimenter in great difficulty, for it subjects him to the temptation to discover "undue similarity" between tests *after* he has found that their correlation breaks the hierarchy. Moreover, whole groups of tests were found to fail to conform; and so group factors were admitted, though always, by the experimenter trained in that school, with reluctance and in as small a number as possible. It had, however, become quite clear that the *Theory* of Two Factors in its original form had been superseded by a theory of many factors, although the *method* of two factors remained as an analytical device for indicating their presence and for isolating them in comparative purity.

Under these circumstances it is not surprising that some workers turned their attention to the possibility of a method of multiple-factor analysis, by which any matrix of test correlations could be analysed direct into its factors (Garnett, 1919*a* and *b*). It was Professor Thurstone of Chicago who saw that one solution to this problem could be reached by a generalization of Spearman's idea of zero tetrad-differences.

2. *Rank of a matrix and number of factors.*—We saw that when all the tetrad differences are zero, the correlations can all be explained by *one* general factor, a tetrad being

formed of the intercorrelations of two tests with two other tests, thus :

	3	4
1	r_{13}	r_{14}
2	r_{23}	r_{24}

and the tetrad-difference being—

$$r_{13}r_{24} - r_{23}r_{14}$$

Thurstone's idea, though rather differently expressed by him, can be based on a second, third, fourth . . . calculation of certain tetrad-differences of tetrad-differences.

To explain this, let us consider the correlation coefficients which three tests make with three others :

	4	5	6
1	r_{14}	r_{15}	r_{16}
2	r_{24}	r_{25}	r_{26}
3	r_{34}	r_{35}	r_{36}

This arrangement of nine correlation coefficients might have been called a " nonad," by analogy with the tetrad. Actually, by mathematicians, it is called a " minor determinant of order three " or more briefly a three-rowed minor ; a tetrad is in this nomenclature a " minor of order two."

We can now, on the above three-rowed determinant, perform the following calculation. Choose the top left coefficient as " pivot," and calculate the four tetrad-differences of which it forms part, namely :

$$(r_{14}r_{25} - r_{24}r_{15}) \qquad (r_{14}r_{26} - r_{24}r_{16})$$
$$(r_{14}r_{35} - r_{34}r_{15}) \qquad (r_{14}r_{36} - r_{34}r_{16})$$

These four tetrad-differences now themselves form a tetrad which can be evaluated. If it is zero, we say that the three-rowed determinant with which we started " vanishes."

Exactly the same repeated process can be carried on with larger minor determinants. For example, the minor of order four here shown vanishes :

	(·26)	·32	·38	·34
	·42	·36	·62	·72
	·44	·62	·66	·46
	·45	·58	·63	·60
for its pivotal t.d.'s are		(— ·0408)	·0016	·0444
		·0204	·0044	— ·0300
		·0068	— ·0072	·0030
and then			(— ·00021216)	·00031824
			·00028288	— ·00042432
and finally			zero	

This process of continually calculating tetrads is called "pivotal condensation." The reader should be given a word of warning here, that the end-result of this form of calculation, if not zero, has to be divided by the product of certain powers of the pivots, to give the value of the determinant we began with. A routine method (Aitken, 1937*a*) of carrying out pivotal condensation, including division by the pivot at each step, is described in Chapter XIV, pages 201ff.*

We can in this way examine the minors of orders two, three, four (and so on) of a correlation matrix, always avoiding those diagonal cells which correspond to the correlation of a test with itself. We may come to a point at which *all* the minors *of that order* vanish. Suppose these minors which all vanish are the minors of order five. We then say that the "rank" of the correlation matrix is four (with the exception of the diagonal cells). There then exists the possibility that the "rank" of the *whole* correlation matrix can be reduced to four by inserting suitable quantities in the diagonal cells (see next section). The "rank" of a matrix is the order of its largest† non-vanish-

* If the process gives, at an earlier stage than the end, a matrix entirely composed of zeros, the rank of the original determinant is correspondingly less, being equal to the number of condensations needed to give zeros.

† "Largest" refers to the number of rows, not to the numerical value.

ing minor. The tests can then be analysed into as many common factors as the above reduced rank of their correlation matrix—the rank, that is to say, apart from the diagonal cells—plus a specific in each test.

3. *Thurstone's method used on a hierarchy.*—Thurstone's rule about the rank includes Spearman's hierarchy as a special case, for in a hierarchy the tetrads—that is, the minors of order *two*—vanish. The rank is therefore *one*, and a hierarchical set of tests can be analysed into *one* common factor plus a specific in each. A simple way of introducing the reader to Thurstone's hypothesis and also to his " centroid " method* of finding a set of factor saturations will be to use it first of all on the perfect Spearman hierarchy which we cited as an artificial example in our first chapter.

Tests	1	2	3	4	5	6
1	.	·72	·63	·54	·45	·36
2	·72	.	·56	·48	·40	·32
3	·63	·56	.	·42	·35	·28
4	·54	·48	·42	.	·30	·24
5	·45	·40	·35	·30	.	·20
6	·36	·32	·28	·24	·20	.

The first step in Thurstone's method, after the rank has been found, is to place in the blank diagonal cells numbers which will cause these cells also to partake of the same rank as the rest of the matrix, numbers which, for a reason which will become clear later, are called " communalities." In our present Spearman example that rank is *one*, i.e. the tetrads vanish. The communalities, therefore, must be such numbers as will make also those tetrads vanish which include a diagonal cell : this enables them to be calculated. Let us, for example, fix our attention on the communality of the first test, which we will designate h_1^2 (the reason for the " square " will become apparent later). Then the tetrad formed by Tests 1 and 2 with Tests 1 and 3 is :

* We shall see why it is called the " centroid " method in the next chapter.

	1	3
1	h_1^2	·63
2	·72	·56

and the tetrad-difference has to vanish. Therefore—

$$\cdot 56 h_1^2 - \cdot 72 \times \cdot 63 = 0$$
$$\therefore h_1^2 = \cdot 81$$

Similarly all the communalities can be calculated, and found to be—

·81 ·64 ·49 ·36 ·25 ·16

(The observant reader will notice that they are the squares of the " saturations " of our first chapter ; but let us continue as though we had not noticed this.)

The method of finding the saturations of each test with the first common factor is then to insert the communalities in the diagonal cells and add up the columns* of the matrix, thus :

Original Correlation Matrix

(·81)	·72	·63	·54	·45	·36	
·72	(·64)	·56	·48	·40	·32	
·63	·56	(·49)	·42	·35	·28	
·54	·48	·42	(·36)	·30	·24	
·45	·40	·35	·30	(·25)	·20	
·36	·32	·28	·24	·20	(·16)	
3·51	3·12	2·73	2·34	1·95	1·56	15·21

The column totals are then themselves added together (15·21) and the square root taken (3·90). The " saturations " of the first (and here the only) common factor are then the columnar totals divided by this square root, namely—

$$\frac{3\cdot 51}{3\cdot 90} \quad \frac{3\cdot 12}{3\cdot 90} \quad \frac{2\cdot 73}{3\cdot 90} \quad \frac{2\cdot 34}{3\cdot 90} \quad \frac{1\cdot 95}{3\cdot 90} \quad \frac{1\cdot 56}{3\cdot 90}$$

or ·9 ·8 ·7 ·6 ·5 ·4

* This, the " centroid " method of finding a set of loadings, is not in any way bound up with Thurstone's theorem about the rank and the number of common factors. It can be used, for example, with unity in each diagonal cell, in which case it will give as many common factors as there are tests, and no specific factors.

as in the present instance we already know them to be. (Very often in multiple-factor analysis the "saturation" of a test with a factor is called the "loading," and this is a convenient place to introduce the new term.)

As applied to the hierarchical case, this method of finding the saturations or loadings had been devised and employed many years previously by Cyril Burt, though it is not quite clear how he would have filled in the blank diagonal cells (Burt, 1917, 53, footnote, and 1940, 448, 462). It should be explained that in actual practice Thurstone and his followers do not calculate the minor determinants to find the rank and the communality, for that would be too laborious. Instead they adopt various approximations, of which the simplest is to insert in each diagonal cell the largest correlation coefficient of the column (see Section 10).

4. *The second stage of the "centroid" method.*—If there is more than one common factor, the process goes on to another stage. Even with our example we can show the beginning of this second stage, which consists in forming that matrix of correlations which the first factor alone would produce. This is done by writing the loadings along the two sides of a chequer board and filling every cell of the chequer board with the product of the loading of that row with the loading of that column, thus :

First-factor Matrix

	·9	·8	·7	·6	·5	·4
·9	·81	·72	·63	·54	·45	·36
·8	·72	·64	·56	·48	·40	·32
·7	·63	·56	·49	·42	·35	·28
·6	·54	·48	·42	·36	·30	·24
·5	·45	·40	·35	·30	·25	·20
·4	·36	·32	·28	·24	·20	·16

This is the "first-factor matrix" which gives the parts of the correlations due to the first factor. This matrix has now to be subtracted from the original matrix to find the residues which must be explained by further common factors.

In our present example the first-factor matrix is identical with the original matrix and *the residues are all zero.* Only the one common factor is therefore required. (Of course, the reader will understand that in a real experimental matrix the residues can never be expected to be *exactly* zero: one is content when they are near enough to zero to be due to chance experimental error.) Had the rank of our original matrix of correlations been, however, higher than *one*, there would have been a matrix of residues.)

Let us now make an artificial example with a larger number of common factors, say *three*, which we can afterwards use to illustrate the further stages of Thurstone's method. We can do this in an illuminating manner by the aid of the oval diagrams described in Chapter I.

5. *A three-factor example.*—In Figure 10, a diagram of the overlapping variances of four tests, let us insert three common factors and specifics to complete the variance of each test to 10 (to make our arithmetical work easy). No factor here is common to all the four tests. The factor with a variance of 4 runs through Tests 1, 2, and 3. That with a variance 3 runs through Tests 2, 3, and 4. That with a variance 2 runs through Tests 1 and 4. The other factors are specifics. The four test variances being each 10, the correlation coefficients are written down from the overlaps by inspection as:

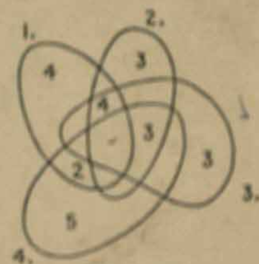


Figure 10.

	1	2	3	4
1	(·6)	·4	·4	·2
2	·4	(·7)	·7	·3
3	·4	·7	(·7)	·3
4	·2	·3	·3	(·5)

Moreover, we can put into our matrix the communalities corresponding to our diagram. Each communality is, in fact, that fraction of the variance of a test which is not specific. Thus ·6 of the variance of Test 1 is "communal," ·4 being specific or "selfish." In this way we have the matrix above, with communalities inserted. We can now pretend that it is an experimental matrix, ready for the application of Thurstone's method, as follows :

	(·6)	·4	·4	·2	
	·4	(·7)	·7	·3	Original
	·4	·7	(·7)	·3	experimental
	·2	·3	·3	(·5)	matrix.
	1·6	2·1	2·1	1·3	$= 7{\cdot}1 = 2{\cdot}6646^2$
1st Loadings	·6005	·7881	·7881	·4879	= 2·6646*
·6005	(·3606)	·4733	·4733	·2930	
·7881	·4733	(·6211)	·6211	·3845	First-factor
·7881	·4733	·6211	(·6211)	·3845	matrix.
·4879	·2930	·3845	·3845	(·2380)	

Here it is seen that the loadings of the first factor, when cross-multiplied in a chequer board, give a first-factor matrix which is *not* identical with the original experimental matrix, unlike the case of the former, hierarchical, matrix. Here (as we who made the matrix know) one factor will not suffice. We subtract the first-factor matrix from the original experimental matrix to see how much of the correlations still has to be explained, and how much of the "communalities" or communal variances. The latter were—

·6 ·7 ·7 ·5

and of these amounts the first factor has explained—

·3606 ·6211 ·6211 ·2380

If we subtract the first-factor matrix, element by element, from the original experimental matrix, we get the residual matrix :

(·2394)	— ·0733	— ·0733	— ·0930	
— ·0733	(·0789)	·0789	— ·0845	First residual
— ·0733	·0789	(·0789)	— ·0845	matrix.
— ·0930	— ·0845	— ·0845	(·2620)	

* This check should always be applied. To avoid complication it is not printed in the later tables. It applies to the loadings with their temporary signs (see page 72).

To this matrix we are now going to apply exactly the same procedure as we applied to the original experimental matrix, in order to find the loadings of the second factor. But we meet at once with a difficulty. The columns of the residual matrix add up exactly* to zero! This always happens, and is indeed a useful check on our arithmetical work up to this point, but it seems to stop our further progress.

To get over this difficulty *we change temporarily the signs of some of the tests* in order to make a majority of the cells of each column of the matrix positive. The best plan is to change the sign of the test with most minuses in its column and row, and so on until there is a large majority of plus signs. Copy the signs on a separate paper, omitting the diagonal signs, which never change. Since some signs will change twice or thrice, use the convention that a plus surrounded by a ring means minus, and if then covered by an X means plus again. Near the end, watch the actual numbers, for the minus signs in a column may be very small. The object is to make the grand total a maximum, and thus take out maximum variance with each factor. We shall here, however, for simplicity adopt an easier rule, i.e. to seek out the column whose total *regardless of signs* is the largest, and then temporarily change the signs of variables so as to make all the signs in that column positive.

The sums of the residual columns, regardless of sign, are—

·4790 ·3156 ·3156 ·5240

and therefore we must change the signs of tests so as to make all the signs in Column 4 positive; that is, we must change the signs of the first three tests.† Since we change the three row signs, as well as the three column signs, this will leave a block of signs unchanged, but will make the last column and the last row all positive. We can then proceed as shown overleaf.

* When enough decimals have been retained. In practice there may be a discrepancy in the last decimal place.

† Changing the sign of Test 4 would here have the same result, but for uniformity of routine we stick to the letter of the rule.

	·2394	− ·0733	− ·0733	(−)·0930	
	− ·0733	·0789	·0789	(−)·0845	First residual
	− ·0733	·0789	·0789	(−)·0845	matrix with
	(−)·0930	(−)·0845	(−)·0845	·2620	changed signs.
	·1858	·1690	·1690	·5240	$= 1{\cdot}0478$
					$= 1{\cdot}0236^2$
2nd Loadings	·1815	·1651	·1651	·5119	With temporary signs.
·1815	·0329	·0300	·0300	·0929	
·1651	·0300	·0273	·0273	·0845	Second-factor
·1651	·0300	·0273	·0273	·0845	matrix.
·5119	·0929	·0845	·0845	·2620	
	·2065	− ·1033	− ·1033	·0001	
	− ·1033	·0516	·0516	.	Second residual
	− ·1033	·0516	·0516	.	matrix.
	·0001	.	.	.	

On the matrix with these temporarily changed signs we then operate exactly as we did on the original experimental matrix, and obtain second-factor loadings which (*with temporary signs*) are—

·1815 ·1651 ·1651 ·5119

The second-factor matrix, that is, the matrix showing how much correlation is due to the second factor, is then made on a chequer board *still using the temporary signs*, and subtracted from the previous matrix of residues (*with its temporary signs*, not with its first signs) to find the residues still remaining, to be explained by further factors. In the present instance we see that the whole variance of the fourth test entirely disappears, and also all the correlations in which that test is concerned.* This test, therefore, is fully explained by the two factors already extracted. Only the first three test variances remain unexhausted, and their correlations. Again the columns of the residual matrix sum exactly to zero. Following our rule, the signs of Tests 2 and 3 have to be temporarily changed before the process can continue. After these changes of sign the

* When enough decimals are retained. We shall treat the ·0001 as zero.

second residual matrix is as follows, and the same operation as before is again performed on it :

	·2065	(−)·1033	(−)·1033	.	Second residual matrix with signs temporarily changed.
	(−)·1033	·0516	·0516	.	
	(−)·1033	·0516	·0516	.	
	.	.	.	.	
	·4131	·2065	·2065	.	$= \cdot 8261 = \cdot 9089^2$
3rd Loadings	·4545	·2272	·2272	.	with temporary signs.

With these third-factor loadings we can now calculate the variances and correlations due to the third factor : and we find these are exactly equal to the second residual matrix. On subtracting, the third residual matrix we obtain is entirely composed of zeros. (In a practical example we should be content if it was sufficiently small.) We thus find (as our construction of the artificial tests entitled us to expect) that the matrix of correlations can be completely explained by three common factors.

After the analysis has been completed, some care is needed in returning from the temporary signs of the loadings to the correct signs. The only safe plan is to write down first of all the loadings with their temporary signs as they came out in the analysis. In our present example these happen to be all positive, though that will not always occur.

Loadings with Temporary Signs

Test	*I*	*II*	*III*
1	·6005	·1815	·4545
2	·7881	·1651	·2272
3	·7881	·1651	·2272
4	·4879	·5119	.

Now, in obtaining Loadings II the signs of Tests 1, 2, and 3 were changed. We must, therefore, in the above table reverse the signs of the loadings of these three tests in Column II *and each later column*. Then in obtaining Loadings III the signs of Test 2 and 3 were changed ; that is, in our case changed back to positive. The loadings with their proper signs are therefore as shown in the first three columns of this table :

Test	*Loadings of the Factors (Signs Replaced)*						
	I	*II*	*III*	*Specific*			
1	·6005	− ·1815	− ·4545	·6324	.	.	.
2	·7881	− ·1651	+ ·2272	.	·5477	.	.
3	·7881	− ·1651	+ ·2272	.	.	·5477	.
4	·4879	·5119	.	.	.	.	·7071

In this table each column of loadings, for the common factors after the first, adds up to zero. The loading of the specific is found from the fact that in each row the sum of the squares must be unity, being the whole variance of the test. The inner product * of each pair of rows gives the correlation between those two tests (Garnett, 1919*a*). Thus—

$$r_{12} = \cdot 6005 \times \cdot 7881 + \cdot 1815 \times \cdot 1651 - \cdot 4545 \times \cdot 2272 = \cdot 4000$$

in agreement with the entry in the original correlation matrix. With artificial data like the present, the analysis results in loadings which give the correlations back exactly.

It will be seen that all the signs in any column of the table of loadings can be reversed without making any change in the inner products of the rows ; that is, without altering the correlations. We would usually prefer, therefore, to reverse the signs of a column like our Column III, so as to make its largest member positive.

The amount which each factor contributes to the variance of the test is indicated by the square of its loading in that test. The sum of the squares of the three common-factor loadings gives the " communality " which we originally

* By the " inner product " of two series of numbers is meant the sum of their products in pairs. Thus the inner product of the two sets :

	a	b	c	d
and	A	B	C	D
is	$aA + bB + cC + dD$			

deduced from Figure 10 and inserted in the diagonal cells of our original correlation matrix. These facts can be better seen if we make a table of the squares of the above loadings :

Test	*Variance contributed by Each Factor*					
	I	*II*	*III*	*Communality*	*Specific Variance*	*Total*
1	·3606	·0329	·2065	·6000	·4000	1
2	·6211	·0273	·0516	·7000	·3000	1
3	·6211	·0273	·0516	·7000	·3000	1
4	·2380	·2620	.	·5000	·5000	1
Total	1·8408	·3495	·3097	2·5000	1·5000	4

6. *Comparison of the analysis with the diagram.*—The reader has probably been turning from this calculation of the factor loadings back to the four-oval diagram with which we started (page 69), to detect any connexion ; and has been disappointed to find none. The fact is that the analysis to which the Thurstone method has led us is, except that it too has three common factors, a different analysis from that which the original diagram naturally invites. That diagram gave for the variance due to each factor the following :

Test	*Variance contributed by Each Factor*					
	I	*II*	*III*	*Communality*	*Specific Variance*	*Total*
1	·4	.	·2	·6	·4	1
2	·4	·3	.	·7	·3	1
3	·4	·3	.	·7	·3	1
4	.	·3	·2	·5	·5	1
Totals	1·2	·9	·4	2·5	1·5	4

and the factor loadings are the positive square roots of these.

Test	Loadings of the Factors						
	I	*II*	*III*	*Specifics*			
1	·6325	.	·4472	·6324	.	.	.
2	·6325	·5477	.	.	·5477	.	.
3	·6325	·5477	.	.	.	·5477	.
4	.	·5477	·4472	.	.	.	·7071

The only points in common between the two analyses are that they both have the same communalities (and therefore the same specific variances) and the same number of common factors. The Thurstone analysis has two general factors (running through all four tests), while the diagram had none: and the Thurstone analysis has several negative loadings, while the diagram had none. We shall see later that Thurstone, after arriving at this first analysis, endeavours to convert it into an analysis more like that of our diagram, with no negative loadings and no completely general factors. This is one of the most difficult yet essential parts of his method.

7. *Analysis into two common factors.*—When we began our analysis of the matrix of correlations corresponding to Figure 10, we simply put the communalities suggested by that figure into the blank diagonal cells. That served to illustrate the fact that the Thurstone method of calculation will bring out as many factors as correspond to the communalities used, here three factors. But it disregarded (intentionally for the purpose of the above illustration) a cardinal point of Thurstone's theory that we must seek for the communalities which make the rank of the matrix a minimum, and therefore the number of common factors a minimum. We simply accepted the communalities suggested by the diagram. Let us now repair our omission and see if there is not a possible analysis of these tests into fewer than three common factors. There is no hope of reducing the rank to one, for the original correlations give two of the three tetrads different from zero, and we may (in an artificial example) assume that there are no experimental or other errors. But there is nothing in the experi-

mental correlations to make it certain that rank 2 cannot be attained. With only four tests (far too few, be it remembered, for an actual experiment) there is no minor of order three entirely composed of experimentally obtained correlations. It may then be the case that communalities can be found which reduce the rank to 2. Indeed, as we shall see presently, many sets of communalities will do so, of which one is shown here :

$$\begin{array}{cccc} (\cdot 2\dot{6}) & \cdot 4 & \cdot 4 & \cdot 2 \\ \cdot 4 & (\cdot 7) & \cdot 7 & \cdot 3 \\ \cdot 4 & \cdot 7 & (\cdot 7) & \cdot 3 \\ \cdot 2 & \cdot 3 & \cdot 3 & (\cdot 15) \end{array}$$

These communalities $\cdot 2\dot{6}$, $\cdot 7$, $\cdot 7$, and $\cdot 15$ make every three-rowed minor exactly zero. For example, the minor

$$\begin{array}{ccc} (\cdot 2\dot{6}) & \cdot 4 & \cdot 2 \\ \cdot 4 & (\cdot 7) & \cdot 3 \\ \cdot 2 & \cdot 3 & (\cdot 15) \end{array}$$

becomes by " pivotal condensation " :

$$\begin{array}{cc} \cdot 02\dot{6} & 0 \\ 0 & 0 \end{array}$$

and finally $\quad 0$

It must, therefore, be possible to make a four-oval diagram, showing only two common factors, and indeed

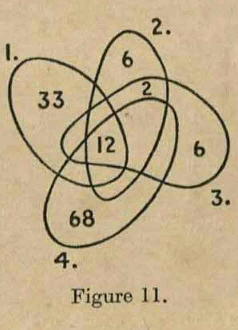


Figure 11.

more than one such diagram can be found. One is shown in Figure 11.

This gives exactly the same correlations. For example—

$$r_{23} = \frac{12 + 2}{\sqrt{(20 \times 20)}} = \frac{14}{20} = \cdot 7$$

$$r_{34} = \frac{12}{\sqrt{(20 \times 80)}} = \frac{12}{40} = \cdot 3$$

It also gives the communalities ·2$\dot{6}$, ·7, ·7, ·15. For example, in Test 1, variance to the amount of 12 out of 45 is communal, and 12/45 = ·2$\dot{6}$.

The insertion of these communalities, therefore, in the matrix of correlations ought to give a matrix which only two applications of Thurstone's calculation should completely exhaust. The reader is advised to carry out the calculation as an exercise. He will find for the first-factor loadings—

·5000 ·8290 ·8290 ·3750

and if in the first residual matrix, following our rule, he changes temporarily the signs of Tests 2 and 3, the second-factor loadings will be—

·1291 − ·1128 − ·1128 ·0968

The second residual matrix will be found to be exactly zero in each of its sixteen cells. The variance (square of the loading) contributed by each factor to each test is then in this analysis :

Test	*Variance contributed by Each Factor*				
	I	*II*	*Communality*	*Specific Variance*	*Total*
1	·2500	·0167	·2667	·7333	1
2	·6873	·0127	·7000	·3000	1
3	·6873	·0127	·7000	·3000	1
4	·1406	·0094	·1500	·8500	1
Totals	1·7652	·0515	1·8167	2·1833	4

If we now compare these analyses, we see that the three common factors of the previous analysis "took out," as the factorial worker says, a variance of 2·5 of the total 4, leaving 1·5 for the specifics. The present analysis leaves 2·1833 for the specifics, which here form a larger part of the four tests.

8. *Alexander's rotation.*—We saw in Section 6 that the Thurstone method there led to an analysis which was different from the analysis corresponding to the diagram with which we began. That is also the case with the present analysis into two common factors—the very fact that it gives the second factor two negative loadings shows this, for the diagram (Figure 11) corresponds to positive loadings only. We said, too, in Section 6 that a difficult part of Thurstone's method was the conversion of the loadings into new and equivalent loadings which are all positive. This will form the subject of a later and more technical chapter ; but a simple illustration of one method of conversion (or "rotation" as it is called, for a reason which will become clear later) can be given from our present example. It is a method which can be used only if we have reason to think that one of our tests contains only one common factor (Alexander, 1935, 144). Let us suppose in our present case that from other sources we know this fact about Test 1. The centroid analysis has given us the loadings shown in the first two columns of this table :

Test	*Unrotated Loadings*		*Communality*	*Rotated Loadings*		*Rotated Loadings*	
	I	*II*		*I**	*II**	*I***	*II***
1	·5000	·1291	·2667	·5164	.	·4781	·1952
2	·8290	− ·1128	·7000	·7746	·3162	·8367	.
3	·8290	− ·1128	·7000	·7746	·3162	·8367	.
4	·3750	·0968	·1500	·3873	.	·3586	·1464

The communalities are also shown ; they are the sums of the squares of the loadings. If now we know or decide to assume that Test 1 has really only one common factor, and if we want to preserve the communalities shown, then the loading of factor I* in Test 1 must be the square root of ·2667, namely ·5164.

The loadings of factor I* in the other three tests can now be found from the fact that they must give the correlations of those tests with Test 1, since Test 1 has no second factor to contribute. The loadings shown in column I* are found in this way: for example, ·7746 is the quotient of ·5164 divided into r_{12} (·4), and ·3873 is similarly r_{14} (·2) divided by ·5164.

The contributions of factor I* to the communalities are obtained by squaring these loadings. In Test 1, we already know that factor I* exhausts the communality, for that is how we found its loading. We discover that in Test 4, factor I* likewise exhausts the communality, for the square of ·3873 is ·1500. The other two tests, however, have each an amount of communality remaining equal to ·1000 (i.e. ·7000 – ·7746²). The square root of ·1000, therefore (·3162), must be the loading of factor II* in Tests 2 and 3. The double column of loadings ought now to give all the correlations of the original correlation matrix, and we find that it does so. Thus, e.g.—

$$r_{23} = \cdot 7746 \times \cdot 7746 + \cdot 3162 \times \cdot 3162 = \cdot 7000$$

$$\text{and} \quad r_{24} = \cdot 7746 \times \cdot 3873 = \cdot 3000$$

Moreover, the analysis into factors I* and II* corresponds exactly to Figure 11. For example, the loading of factor II* in Test 2 in that diagram is the square root of 2/20 (·3162); and the loading of factor I* in Test 4 is the square root of 12/80 (·3873).

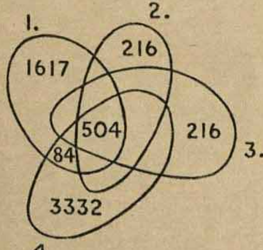


Figure 12.

If, however, the experimenter had reasons for thinking that Test 2 (not Test 1) was free from the second common factor, his " rotation " of the loadings would have given a different result, shown in the table on page 79 in column I** and II**. This set of loadings also gives the correct communalities and the experimental correlations, but does not correspond to Figure 11. A diagram can, however, be constructed to agree with it (Figure 12) and the reader is advised to check the agreement by calculating from the diagram the loadings of each factor, the communalities of each test, and the correlations.

We have had, in Figures 10, 11, and 12, three different analyses of the same matrix of correlations. If with Thurstone we decide that analyses must always use the minimal number of common factors, we will reject Figure 10. Between Figures 11 and 12, however, this principle makes no choice. Much of the later and more technical part of Thurstone's method is taken up with his endeavours to lay down conditions which will make the analysis unique.

9. *Unique communalities.*—The first requirement for a unique analysis is that the set of communalities which gives the lowest rank should be unique, and this is not the case with a battery of only four tests and minimal rank 2, like our example. There are many different sets of communalities, all of which reduce the matrix of correlations of our four tests to rank 2. If, for example, we fix the first communality arbitrarily, say at ·5, we can condense the determinant to one of order 3 by using ·5 as a pivot (as on page 65) except that the diagonal of the smaller matrix will be blank :

(·5)	·4	·4	·2
·4	.	·7	·3
·4	·7	.	·3
·2	·3	·3	.
	.	·19	·07
	·19	.	·07
	·07	·07	.

We can then fill the diagonal of the smaller matrix with numbers which will make each of its tetrads zero, namely—

·19 ·19 ·0258

and then, working back to the original matrix, find the communalities—

·5 ·7 ·7 ·1316

which make its rank exactly 2. We can similarly insert different numbers for the first communality and calculate different sets of communalities, any one set of which will reduce the rank to 2. In this way we can go from 1·0 down to 0·22951 for the first communality without obtaining inadmissible magnitudes for the others. Some sets are given in the following table * :

1	2	3	4	*Sum*
1·0	·7	·7	·12963	2·52963
·5	·7	·7	·13158	2·03158
·3	·7	·7	·14	1·84
·2$\dot{6}$	·7	·7	·15	1·81$\dot{6}$
·25	·7	·7	·1$\dot{6}$	1·81$\dot{6}$
·24	·7	·7	·20	1·84
·22951	·7	·7	1·0	2·62951

If, however, we search for and find a fifth test to add to the four, which will still permit the rank to be reduced to 2, this fifth test will fix the communalities at some point or other within the above range. Suppose that this test gave the correlations shown in the last row and column :

	1	2	3	4	5
1	.	·4	·4	·2	·5883
2	·4	.	·7	·3	·2852
3	·4	·7	.	·3	·2852
4	·2	·3	·3	.	·1480
5	·5883	·2852	·2852	·1480	.

If we now try to find communalities to reduce this matrix to rank 2 (as can be done), we find only the one set—

·7 ·7 ·7 ·13030 ·5

The reader can try this by assigning an arbitrary value for the first one,† and then condensing the matrix on the lines

* The circumstance that the communalities of Tests 2 and 3 remain fixed and alike is due to these tests being identical except for their specific. This lightens the arithmetic, but would not occur in practice.

† Alternatively, the communalities (which are now unique) can be found by equating to zero those three-rowed minors which have only one element in common with the diagonal. In this connexion see Ledermann, 1937*a*.

employed above, when he will always find some obstacle in the way unless he chooses ·7. Try, for example, ·5 for the first communality :

(·5)	·4	·4	·2	·5883
·4	.	·7	·3	·2852
·4	·7	.	·3	·2852
·2	·3	·3	.	·1480
·5883	·2852	·2852	·1480	.
	(x)	·19	·07	— ·09272
	·19	.	·07	— ·09272
	·07	·07	.	— ·04366
	— ·09272	— ·09272	— ·04366	.

Now, if the upper matrix is to be of rank 2, the second condensation must give only zeros (see footnote, page 65). But if we fix our attention on different tetrads in the lower matrix which contain the pivot x, we see that they give, if they have to be zero, incompatible values for x. Thus from one tetrad we get $x = \cdot 19$, from another $x = \cdot 14866$. With ·5 as first communality, rank 2 cannot be attained. With five tests (or more), if rank 2 can be attained at all, it can be by only one unique set of communalities. Just as it took three tests to enable the saturations with Spearman's g to be calculated, so it takes five tests to enable communalities due to two common factors to be calculated. For larger numbers of common factors, the number of tests required to make the set of communalities unique is shown in the following table (*Vectors*, 77). The lower numbers * are given by the formula—

$$n \geqslant \frac{(2r + 1) + \sqrt{(8r + 1)}}{2}$$

r Factors	1	2	3	4	5	6	7	8	9	10	11	12
n Tests	3	5	6	8	9	10	12	13	14	15	17	18

* With six tests the communalities which reduce to rank 3 are not necessarily unique, for there are, or there may be, two sets of them. See Wilson and Worcester, 1939.

I think the ambiguity, which is not practically important, only occurs when n is exactly equal to the formula, e.g. when $r = 3$, 6, 10, etc.

If we were actually confronted with the matrix of correlations shown on page 69, and asked what the communalities were which reduced it to the lowest possible rank, we would find it very unsatisfactory to have to guess at random and try each set ; and our embarrassment would be still greater if there were more tests in the battery, as would actually be the case in practice. There would also be sampling error (which in this our preliminary description of Thurstone's method we are assuming to be non-existent). Under these circumstances, devices for arriving rapidly at approximate values of the communalities are very desirable.

✓ 10. *Method of approximating to the communalities.*—Thurstone has described many ways of estimating the communalities, and articles still issue from his laboratory on this subject. (He points out, however, that if the number of tests is fairly large, an exact estimate is not very important, and can in any case be improved by iteration, using the sums of squares of the loadings for a new estimate.

The simplest plan is to use as an approximate communality the largest correlation coefficient in the column. That this is plausible can be seen from a consideration of the case where there is only one factor, when the communality of Test 1 would be $r_{12} \cdot r_{13}/r_{23}$, which is likely to be roughly equal to either r_{12} or r_{13} if these tests correlate highly with Test 1 and probably therefore with each other.

We shall illustrate this, the easiest, method on the same example as we used above, for the sake of comparison and for ease in arithmetical computation, even although that example is really an exact and artificial one unclouded by sampling error. Inserting then the highest coefficients in each column we get :

(·5883)	·4	·4	·2	·5883	
·4	(·7)	·7	·3	·2852	
·4	·7	(·7)	·3	·2852	
·2	·3	·3	(·3)	·1480	
·5883	·2852	·2852	·1480	(·5883)	
2·1766	2·3852	2·3852	1·2480	1·8950	= 10·0900
					= $3{\cdot}1765^2$

First Loadings	·6852	·7509	·7509	·3929	·5966

The communalities which really give the minimum rank are, as we saw on page 82—

·7 ·7 ·7 ·1303 ·5

and the correct first-factor loadings obtained by their use—

·7257 ·7564 ·7564 ·3420 ·5729

With a large battery the difference between the loadings obtained by the approximation and by the correct communalities would be much less. For the " centroid " method depends on the *relative* totals of the columns of the correlation matrix ; and when there are twenty or more tests, these relative totals will not be seriously changed by the exact value given to the communality in the column. When the number of tests is large, the influence of the one communality in each column is swamped by the influence of the numerous correlations.

The process now goes on as on page 71, and the residuals left after subtraction of the first-factor matrix check by summing in each column to zero, as there.

Before, however, proceeding any farther, in this approximate method *we delete the quantities in the diagonal* (the residues of the guessed communalities) *and replace them by the largest coefficient in the column* regardless of its sign, which we change to plus in the diagonal cell if it is negative in its own cell.* The reason for this is apparent, especially when, as may and does happen, the existing diagonal residues are negative, which is theoretically impossible. For although the guessing of the first communalities does not in a large battery make much difference to the first-factor loadings, it may make a big difference to the diagonal residues. If the battery is very large indeed, our first-factor loadings would come out much the same, even if we entered *zero* for every communality, but the diagonal residues would then all be negative. In short, the diagonal residues are much the least trustworthy part of the calcu-

* It is necessary to keep an eye on the fact that what is inserted must not, with the squares of the previous loadings of that test, amount to more than unity.

lation when approximate communalities are used, and it is better to delete them at each stage and make a new approximation.

11. *Illustrated on the example.*—To make this clearer, the whole approximate process is here set out for our small example as far as the second residual matrix. The explanations printed alongside the calculation will make each stage clear. It is important to form the residual matrices exactly as instructed, as otherwise the check of the columns summing to zero will not work. In practice, certainly if a calculating machine were being used, several of the matrices here printed for clearness would be omitted ; for example, with a machine one would go straight from *A* to *C*, while *D* and *E* would be made by actually altering *C* itself :

A		(·5883)	·4	·4	·2	·5883	Largest *r* of column inserted in diagonal cell.
		·4	(·7)	·7	·3	·2852	
		·4	·7	(·7)	·3	·2852	
		·2	·3	·3	(·3)	·1480	
		·5883	·2852	·2852	·1480	(·5883)	
		2·1766	2·3852	2·3852	1·2480	1·8950	= 10·0900
							= 3·1765²
Loadings I		·6852	·7509	·7509	·3929	·5966	= 3·1765
B	·6852	(·4695)	·5145	·5145	·2692	·4088	First-factor matrix.
	·7509	·5145	(·5639)	·5639	·2950	·4480	
	·7509	·5145	·5639	(·5639)	·2950	·4480	
	·3929	·2692	·2950	·2950	(·1544)	·2344	
	·5966	·4088	·4480	·4480	·2344	(·3559)	
C		(·1188)	−·1145	−·1145	−·0692	·1795	First residual matrix. *A* − *B*
		−·1145	(·1361)	·1361	·0050	−·1628	
		−·1145	·1361	(·1361)	·0050	−·1628	
		−·0692	·0050	·0050	(·1456)	−·0864	
		·1795	−·1628	−·1628	−·0864	(·2324)	
		·0001	−·0001	−·0001	·0000	−·0001	Columns check to zero.
D		(·1795)	−·1145	−·1145	−·0692	·1795	Largest *r* of each column (regardless of sign) inserted in each diagonal cell.
		−·1145	(·1628)	·1361	·0050	−·1628	
		−·1145	·1361	(·1628)	·0050	−·1628	
		−·0692	·0050	·0050	(·0864)	−·0864	
		·1795	−·1628	−·1628	−·0864	(·1795)	

		·6572	·5812	·5812	·2520	·7710	Sum disregarding signs.
E		(·1795)	·1145	·1145	·0692	·1795	Signs of Tests 2, 3, and 4 changed to make largest column (·7710) all positive.
		·1145	(·1628)	·1361	·0050	·1628	
		·1145	·1361	(·1628)	·0050	·1628	
		·0692	·0050	·0050	(·0864)	·0864	
		·1795	·1628	·1628	·0864	(·1795)	
Algebraic Sum		·6572	·5812	·5812	·2520	·7710	$= 2{\cdot}8426$ $= 1{\cdot}6860^2$
Loadings II		·3898	·3447	·3447	·1495	·4573	(With temporary signs.)
F	·3898	(·1519)	·1344	·1344	·0583	·1783	Second-factor matrix, using temporary signs.
	·3447	·1344	(·1188)	·1188	·0515	·1576	
	·3447	·1344	·1188	(·1188)	·0515	·1576	
	·1495	·0583	·0515	·0515	(·0124)	·0683	
	·4573	·1783	·1576	·1576	·0683	(·2091)	
G		(·0276)	−·0199	−·0199	·0109	·0012	Second residual matrix. $E - F$
		−·0199	(·0440)	·0173	−·0465	·0052	
		−·0199	·0173	(·0440)	−·0465	·0052	
		·0109	−·0465	−·0465	(·0640)	·0180	
		·0012	·0052	·0052	·0180	(−·0296)	
		−·0001	−·0001	·0001	−·0001	·0000	Columns check to zero.

Notes.—It is fortuitous that *all* the entries in *E* are positive. Usually some will be negative.

In the check for the residual matrices, a discrepancy from zero in the last figure is often to be expected, even of three or four units in a large matrix.

Note the negative value occurring in a diagonal cell in *G*.

Further stages would be carried on in the same way. But at each stage the residues will be examined to see if further analysis is worth while, by methods indicated later. Meanwhile let us assume in the present example that no more factors need be extracted.

The matrix of loadings of common factors thus arrived at is, after we have replaced the proper signs in Loadings II, shown at the top of the next page.

The communalities ·6214, etc., are the sums of the squares of the two loadings. For comparison with the

Test	*Approximate Method*			*True Values*
	I	*II*	*Communality*	*Communality*
1	·6852	·3898	·6214	·7000
2	·7509	− ·3447	·6827	·7000
3	·7509	− ·3447	·6827	·7000
4	·3929	− ·1495	·1767	·1303
5	·5966	·4573	·5651	·5000
			2·7286	2·7303

approximate communalities thus obtained there are shown the true values, which in this artificial case are known to us (see Section 9). This is for instructional purposes only—the comparison is not intended as any criticism of Thurstone's method of approximation. As has been explained, this method is used only on large batteries, and it is a very severe test indeed to employ it on a battery of only five tests.

√ 12. *Iteration of the process to improve the communalities.*—We might now go back and begin our whole calculation again, using the communalities ·6214, etc., arrived at by the first approximation. This does not seem often to be done in practice, most workers being content with the approximation first arrived at. If we repeat the calculation again and again with our present example, on each occasion using as communalities the sum of the squares of the loadings given by the preceding calculation, we get the following sets of closer and closer approximation to the true communalities* :

	h_1^2	h_2^2	h_3^2	h_4^2	h_5^2
First trial communalities	·5883	·7000	·7000	·3000	·5883
Next approximation	·6214	·6827	·6827	·1767	·5651
Next approximation	·6381	·6970	·6970	·1477	·5392
Next approximation	·6535	·7043	·7043	·1397	·5253
True values	·7000	·7000	·7000	·1303	·5000

* In these repetitions we do not, as in the case of the first guess, alter the diagonal cells in each matrix of residues : we retain the diagonal residues without change.

The example has served to show how to work the iterative method of approximating to the communalities. Being an artificial example, and not overlaid with sampling error, it has had the advantage of allowing us to compare the approximations with the true values. But it must be remembered that a real experimental matrix is not likely to have an *exact* low rank to which approximation can converge as here. In that case the approximations will presumably give an indication of the low rank which the matrix *nearly* has, which it might be made to have by small adjustments in its elements.

It should be pointed out that iteration of each factor extraction separately will not give the same result. By iteration of the factors one by one we mean that after the loadings of the first factor are obtained they are squared and put into the diagonal cells as new communalities, and this is repeated again and again until the communalities remain unchanged. When this point is reached, the original matrix of correlations has been reduced as nearly to rank *one* as is possible.

If the residues, after removal of the first factor, are then (after sign-changing) treated in the same way, they in turn will be reduced as nearly as possible to rank *one*. And so with successive residues, each matrix of residues being in succession reduced as nearly as possible to rank *one* by iteration of the one summation only. This process, although much easier than reiterating the whole process, and to that extent excusable, will not give the lowest possible rank for the whole. Consider, for example, the correlations of the five tests used above on page 82. When communalities are reiterated with the first factor only, they settle down rapidly (the reader should check this) to—

·4571 ·5421 ·5421 ·1261 ·2729

When the residues then left are taken, and a factor taken out and iterated, the communalities settle down to—

·1677 ·1003 ·1003 ·0113 ·1680

The sum of these first-factor and second-factor sets is the set—

·6248 ·6424 ·6424 ·1374 ·4409

These, however, if inserted in the diagonal cells of the original matrix, do not reduce it exactly to rank *two*, as can be done by the true communalities—

·7000 ·7000 ·7000 ·1303 ·5000

Iteration over two factors, as shown in the table on page 88, produces with four repetitions the approximations—

·6535 ·7043 ·7043 ·1397 ·5253

and (since in this artificial example rank *two* can be exactly reached) would ultimately converge to the above true values, though at the expense of much labour, for the convergence is slow. The iteration of each factor separately, however, would *never* converge to the true values. The above values (·6248, etc.) are final, and yet do not give rank *two*.

13. *Other methods of assessing the communalities.*—The labour of finding the minimum communalities by iteration is so great that methods of improving the first guess are desirable. Medland (*Pmka.* 1947, **12**, 101–10) has tried nine such methods on a correlation matrix with 63 variables. A method entitled *Centroid No.* 1 *method* seemed to be best. A sub-group is chosen of from three to five tests which correlate most highly with the test whose communality is wanted. The highest correlation t in each column of the sub-group is inserted in the diagonal cell, and the columns summed. The grand total is also found. Then the estimate of h_1^2 is—

$$\frac{(\Sigma r_1 + t_1)^2}{\Sigma r + \Sigma t}$$

where the numerator is the square of the column total, and the denominator is the grand total. Thus if the correlations of the sub-group were—

(·72)	·72	·63	·24	
·72	(·72)	·47	·59	
·63	·47	(·63)	·41	
·24	·59	·41	(·59)	
2·31	2·50	2·14	1·83	= 8·78

the estimate of h_1^2 would be—

$$\frac{2{\cdot}31^2}{8{\cdot}78} = {\cdot}608$$

Clearly the same sub-group will usually serve for more than one of its members. Thus from the above example h_2^2 can be estimated to be ·712.

A graphical method, for which the reader is referred to Medland's article, was about equally accurate but more laborious. Rosner (*Pmka.* 1948, **13,** 181–4) gives an algebraic solution for the communalities depending upon the Cayley-Hamilton theorem that any square matrix satisfies its own characteristic equation, but adds that the method "is not at all suited for practical purposes. The computational labour is prohibitive." It is, however, interesting theoretically and may suggest new advances.

CHAPTER VI

THE GEOMETRICAL PICTURE

1. *The scatter-diagram of two tests.*—A well-known way of representing correlation, and that used by Sir Francis Galton who devised correlation coefficients, is by a scatter-diagram. The scores in two tests are used as rectangular abscissæ and ordinates, and each person represented by a

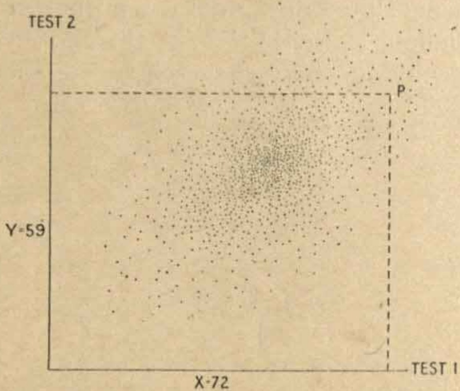


Figure 13.

dot. Thus, if a person makes a score of $X = 72$ in a Test 1 and of $Y = 59$ in a Test 2, he is represented by the point P. The two tests are represented by the rectangular axes. If a large number of persons take the two tests, their points form the "scatter-diagram," looking like a lot of shots at a target. The dots are most densely crowded together near a point whose ordinates are the average scores in the two tests. If there is no correlation between the two tests, and suitable units are used, the dots will thin out equally in all directions, forming a circular-shaped group. If, on the other hand, there is correlation, the group of dots will be elliptical in appearance, with an axis slanting-wise inclined to the test lines ; and more and more elliptical—

the closer the resemblance of the scores, the higher, that is, the correlation. If we have first standardized the scores, the test lines will pass through the centre of the group, the average, and the axis of the ellipse will be equally inclined to both tests. In Figure 14 it is indicated how the elliptical group of dots narrows in the one direction, and lengthens in the other, with increasing correlation. The circle corresponds to zero correlation, the fat ellipse to

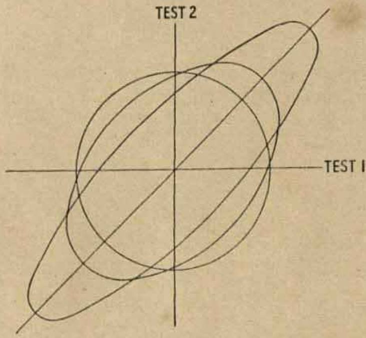


Figure 14.

$r = \cdot 5$, the long thin one to $r = \cdot 9$. In perfect correlation all the dots would be on a line. In negative correlation the ellipse would be slanting the other way. These ellipses must not be looked upon as bounding the group of dots, which thins out to an indefinite distance. They are like contours of a hill, being, in fact, "contours" of the density of the dots.

2. *Three tests.*—When we have three tests we need three rectangular axes, like the three lines which meet in the corner of a room. A person's three scores, measured along these lines, define a point in solid space, a point in the room. The points thus representing a large number of persons will form a swarm in the room, congregated most thickly round the man who is average in all three tests, like a swarm of bees round the queen. If there is no correlation between any of the tests and suitable units are used, the swarm will be globular, but if there is correlation it will lengthen into an ellipsoidal shape like a Rugby football or a Zeppelin, though its waistline need not be

circular. In place of the ellipses of the two-dimensional figure, we now have ellipsoidal shells of equal density of the dots representing persons. One such is shown in Figure 15, which the reader can imagine as being the room in which he is seated, the test lines, in their positive halves, being represented by the three edges of floor and walls

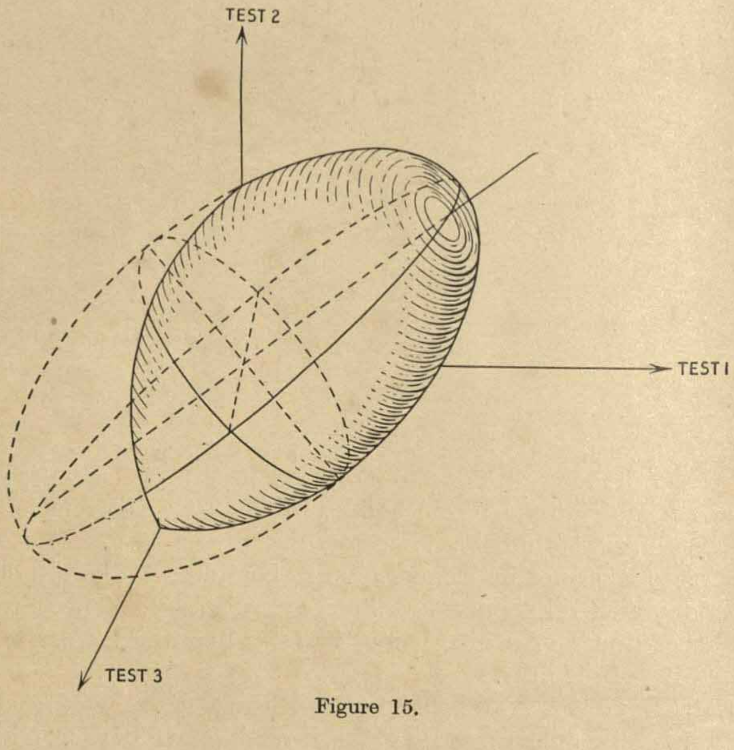
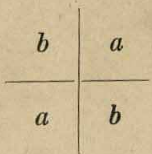


Figure 15.

which meet in a corner, where the point representing the average man is placed. The ellipsoidal swarm is then partly in the room, partly outside and below it. The part of the swarm in the room (in the positive octant, that is) is composed of persons scoring above the average in all three tests. The end of the major axis of the ellipsoid, that is, the longest line that can be drawn in it, is shown projecting. If the tests have all been standardized, this major axis will be equally inclined to their lines. The shadow of the ellipsoid, projected at right angles on to a wall or the floor of the room, will be a correlational ellipse due to the two tests edging that wall, or edging the floor. These three silhouettes will in general be different, depending on the adiposity, as it were, of the ellipsoid.

When we have more than three tests we cannot make or easily imagine a similar model, for we know in real life only space of three dimensions. But mathematically we still can conceive of as many rectangular axes as there are tests, in a " space " of more dimensions, of as many dimensions, indeed, as the number of tests. And we still speak of the " ellipsoidal " shape of the swarm of persons.

3. *The four quadrants.*—Let us now return to the case of two tests. If the persons tested are numerous it will,



with most tests, be found that the numbers in the two quadrants marked *a* are approximately equal (the axes being drawn, it is understood, through the average score of each test) and, similarly, the numbers in the two quadrants marked *b* in the figure.

A portion *a* of the crowd of persons, that is, get scores above the average in each test, and an equal portion *a* are below the average in each. These people add to the correlation between the tests, whereas the others, in the *b* quadrants, are all good in one but bad in the other test

d detrac

ERRATA (Fifth Edition)

The last complete sentence on page 94 and the last sentence of section 5 on page 97 are incorrect and should be deleted. The major axis is *not* equally inclined, in general, to the orthogonal test lines.

then

$$r = \cos \theta = \cos \frac{1000}{3000} \times 180° = \cos 60° = 0{\cdot}5$$

Actual correlation tables will, of course, not show such complete equality in the opposite quadrants, and, moreover, the reader must beware of applying this formula unless the dividing lines are drawn through the means.

4. *Making the crowd circular.*—We are next going to make a change in our model by rotating the two test vectors, hitherto at right angles, towards one another until the angle between them is the above angle θ, whose cosine is the correlation coefficient. A person's point *P* will still be located at the point where the two perpendiculars from his scores meet. The rotation of the test lines towards one another, pivoted on the average man at the point where they cross, will, however, move the dots representing persons, and move them in such a way that the elliptical shape of the crowd disappears and it becomes circular. The presence of correlation is not now shown by the configuration of the crowd, but by the angle between the test lines. The cosine of this angle is the correlation coefficient.

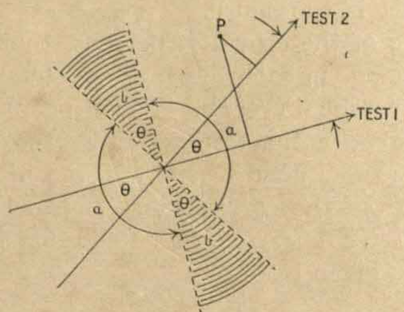


Figure 16.

If we guide the eye by drawing a dotted line at right angles to each test line, we see that our former quadrants *a* and *b* are now represented by sectors of the circular crowd. Perpendiculars from any point in the white sectors *a* on to the test lines both fall on the same side of the average: all persons situated in these sectors are either above the average in both tests (like P) or below in both. Anyone, on the other hand, whose point is in the shaded sector b is above the average in one of the tests and below in the other. Those in a add to the correlation, those in b diminish it. If correlation is perfect, the two test lines must be brought together until they coincide: and then the dotted lines will also coincide and the sector b will disappear. If, on the other hand, the correlation is low, the test lines will have to be farther apart, and the sector b will increase, until, when correlation is zero, the test lines are at right angles and the sectors a and b are equal and balance one another, the *pros* equal to the *cons*. For negative correlation the angle θ between the test lines becomes obtuse, and the sectors b larger than the sectors a.

5. *Ellipsoid into sphere.*—With three tests we saw that the solid "scatter-diagram," made with the test lines at right angles to one another, was ellipsoidal in form. Just as we converted the elliptical two-dimensional scatter-diagram into a circular crowd of dots by bringing the test lines closer together, until the cosine of the angle between them equalled the correlation coefficient, so with the ellipsoidal swarm of dots when we have three tests. If we take hold of the three test lines and swivel them nearer to each other, until the angle between each pair represents their correlation coefficient by its cosine, we then find that the ellipsoid has become a sphere. Moreover, we then find that the long major axis of the ellipsoid, which (with standardized tests) was equally inclined to the three rectangular test lines, is not now equally inclined to them, now that they have been brought into their new positions—unless, indeed, all three correlations are exactly equal.

6. *A wire model.*—Let us suppose we want to make a wire model of this arrangement of three test lines, supposing that we have calculated by the usual product-moment formula the three correlation coefficients. Choosing any two of the tests, we find from a table of cosines what angle has a cosine equal to their correlation coefficient, and we lay two straight wires on the table crossing one another at

this angle, like an X. Imagine them soldered together at the point where they cross, which represents the man average in each test.

Now consider the third test, and look up the angles whose cosines equal its correlation coefficients with Tests 1 and 2. The wire for this third test must be so placed as to make these angles with the first two wires—and we find that it will not lie flat on the table but sticks up at an angle, and its negative half has to go through the table and stick out below it. If we solder the three wires together where they cross (at the point representing the man who gets the average score in each of the three tests) and pick them up, they form a double tripod.

7. *Two kinds of space.*—It will be seen that we have described two geometrical ways of representing correlation using two different spaces. In the one kind of space, the test lines are at right angles to one another, or orthogonal, and the presence of correlation is shown by the fact that the swarm of dots representing persons is not spherical but ellipsoidal.

In the other kind of space, the crowd of dots representing persons is spherical and the presence of correlation is shown by the test lines *not* being orthogonal but at angles with one another whose cosines equal the correlation coefficients.

In both kinds of space, a person's scores in the tests are found by dropping perpendiculars from his point on to the test lines. The distances of the feet of these perpendiculars from the origin—that is, from the point where the test lines cross—are his scores in the tests.

If the test lines in this second kind of space are swivelled back into orthogonality, the person-points will move, will cease to be spherical in contour, and become ellipsoidal. All this is true, not only for three-dimensional space, when we have only three tests, but for multi-dimensional space needed to represent many tests and their inter-correlations. The algebra is exactly the same for any number of dimensions, and we continue, in the larger spaces, to use by analogy the terms we are accustomed to in real space, such as sphere, ellipsoid, etc.

8. *A still larger space.*—Another way of arriving at the second of the above two kinds of space—the spherical one, in which the cosines equal the correlation coefficients—is to begin with a much larger space, of as many dimensions as there are persons, who are therein represented by orthogonal axes. If along each person's axis we set off the score he gets in a given test, say Test 1, these abscissæ will define a point in the space representing that test. In the same way each test can be represented by a point. It is a scatter-diagram with the usual rôles of tests and persons exchanged.

These test points will usually be much less numerous than the persons, and they define a sub-space of dimensions equal to the number of tests. This sub-space, if the test scores have been normalized,* is the same as our spherical space, and the lines joining the origin to the test points are our former lines, separated by angles whose cosines equal the correlation coefficients.

9. *Factor axes.*—The problem of factorial analysis is to decide upon a set of axes to use in the space in which the test lines exist. Let us explain this first of all in the simplest case, that of two tests, represented by their lines in a plane, at the angle corresponding to their correlation.

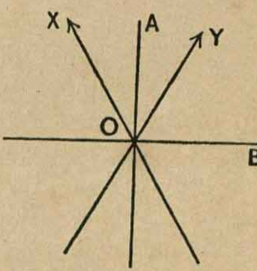


Figure 17.

In this case, the most natural way of drawing orthogonal axes on the paper is to place one of them (see Figure 17) half-way between the test vectors, and the other, of course, at right angles to the first. Of these two factor axes, OA is as near as it can be to both test lines.

We pictured, before, a swarm of ten thousand dots on the paper, each representing a person by his scores in the two tests, found by dropping perpendiculars from his dot to the two vectors. Instead of describing each point (each person, that is) by the two test scores, it is clear that we could describe it by the two factor scores—the feet of

* See footnote, page 6.

perpendiculars on to the factor axes. It is also clear that, as far as this purpose goes, we might have taken our factor axes anywhere, and not necessarily in the positions OA and OB, provided they went through the point O and were at right angles. In other words, we can " rotate " OA and OB round the point O, and any position is equally good for describing the crowd of persons. Either of the tests, indeed, might be made one of the factors. The positions shown in Figure 17 are advantageous only if we want to use only one of our factors and discard the other, in which case obviously OA is the one to keep, as it lies as near as possible to both test axes. The scores along OA are the best possible single description of the two test results.

10. *Spearman axes for two tests.*—The orthogonal axes chosen by Spearman for his factors are, however, none of the positions to which OA and OB can be rotated in the plane of the paper. Besides, Spearman has three factors, and therefore three axes, for two tests, namely the general factor and the two specific factors, and we cannot have three orthogonal axes or factor vectors on a sheet of paper. The Spearman factors must, for two tests, lie in three-dimensional space, like the three lines which meet in the corner of a room. If we rotate the OA and OB of Figure 17 out of the plane of the paper (say, pushing A below the surface of the paper, and, say, raising B above it), we shall clearly have to add a third axis, at right angles to OA and OB, to enable us to describe the tests and the persons who remain on the paper. There are now three axes to rotate; and they must rotate rigidly, remaining at right angles to one another. The point at which Spearman stops the rotation, and decides that the lines then represent the " best " factors, is a position in which one of the axes is at right angles to Test X, and another is at right angles to Test Y. The third axis then represents g.

11. *Spearman axes for four tests.*—We are accustomed to depicting three dimensions on a flat sheet of paper, and so we can, in Figure 18, represent the Spearman axes g, s_1, and s_2 for two tests. And since we have begun to depict other dimensions, by means of perspective, on a flat sheet,

let us continue the process and by a kind of super-perspective imagine that the lines s_3, s_4, and any others we may care to add, represent axes sticking out into a fourth, a fifth, and higher dimensions. Figure 18 thus represents the five Spearman axes for four tests, of which only the line of the first test is shown (in its positive half only).

All the five lines g, s_1, s_2, s_3, and s_4 must be imagined as being each at right angles to all the others in five-dimensional space. The line of Test 1, shown in the diagram, lies *in* the plane or wall edged by g and s_1. It forms acute angles with g and with s_1, the cosines of which angles are its saturations with g and s_1 respectively. If it had been highly saturated with g, it would have leaned nearer to g and farther away from s_1.

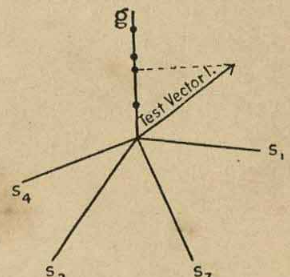


Figure 18.

The other three axes, s_2, s_3, and s_4, are all at right angles to the wall or plane in which Test 1 lies. They have, therefore, no correlation with Test 1, no share in its composition. Test line 2 similarly lies in the wall edged by g and s_2, test line 3 in that edged by g and s_3. The axis g forms a common edge to all these planes. If the battery of tests is hierarchical—that is, if the tetrad-differences are all zero—then all the tests of the battery can be depicted in this way, each in its own plane at right angles to all the other planes, no test line being in the spaces between the " walls."

The four test lines themselves, of course, are only in a four-dimensional space (a 4-space we shall say, for brevity). Just as, when we were discussing Figure 17, we said that Spearman used three axes which were all out of the plane of the paper, so here in Figure 18, with four test lines (only one shown) in a 4-space, Spearman uses five axes in a space of one dimension higher than the number of tests. For n hierarchical tests, Spearman's factors are in an $(n + 1)$-space.

If along each test line we measure the same distance as a unit, then perpendiculars from these points* on to the g axis will give the saturations of the tests with g as fractions of this unit distance. The four dots on the g axis in Figure 18 may thus be taken as representing the test vectors † projected on to the " common-factor space," which is here a line, a space of one dimension only. Thurstone's system is like Spearman's except that the common-factor space is of more dimensions, as many as there are common factors. Figure 19 shows the Thurstone axes for four tests whose matrix of correlation coefficients can be reduced to rank 2.

12. *A common-factor space of two dimensions.*—Here there are two common factors, a and b, and four specifics, s_1, s_2, s_3, and s_4. All the six axes representing these factors in the figure are to be imagined as existing in a 6-space, each at right angles to all the others. The common-factor space is here two-dimensional, the plane or wall edged by a and b—to make it stand out in the figure, a door and a window have been sketched upon it.

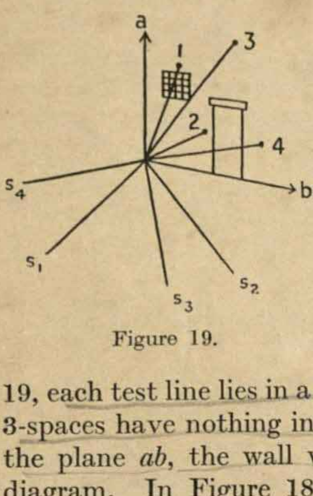


Figure 19.

In Spearman's Figure 18, each test line lay in a plane defined by g and one of the specific axes. Here in Figure 19, each test line lies in a different 3-space. These different 3-spaces have nothing in common with one another except the plane ab, the wall with the door and window in the diagram. In Figure 18 the projections of the unit test vectors on to the common-factor space were lines which all coincided in direction (though they were of different lengths), for there the common-factor space was a line. Here the common-factor space is a plane, and the projections of the four test vectors on to that plane are shown

* These points are then the same as those arrived at by the process described in Section 8 (page 99).

† A vector is a direction with a magnitude, and now that we have measured unit distance along each test line, we may speak of unit test *vectors*.

in the figure by the numbered lines on the "wall." These lines, if they are all projections of vectors of unit length, will by their lengths on the wall represent the square roots of the communalities.

13. *The common-factor space in general.*—When there are r common factors, the common-factor space is of r dimensions, and the whole factor space (including the specifics) is of $(n + r)$ dimensions. The test vectors themselves are in an n-space; their projections on to the common-factor space are crowded into an r-space, and are naturally at smaller angles with one another than the actual test vectors are. These angles between the *projected* test vectors do not, therefore, represent by their cosines the correlations between the tests. The angles are too small for that, and the cosines, therefore, too large. But if we multiply the cosine of such an angle by the lengths of the two projections which it lies between, we again arrive at the correlation.

Thus in Figure 19, the angle between the lines 1 and 3 on the wall is less than the angle between the actual test vectors 1 and 3 out in the 6-space, of which the lines on the wall are the projections. But the lengths of the lines 1 and 3 on the wall are less than the unit length we marked off on the actual vectors, being, in fact, the roots of the communalities. If we call these lengths on the wall h_1 and h_3, then the product h_1h_3 times the cosine of the projected angle again gives the correlation coefficient.

14. *Rotations.*—It will be remembered that Thurstone, after obtaining a set of loadings for the common factors by his method of analysis of the matrix of correlations, "rotates" the axes until the loadings are all positive—and he also likes to make as many of them as possible zero. It is instructive to look at this procedure in the light of our geometrical picture from which the phrase "rotating the factors" is taken. It should be emphasized first of all that such rotation of the common-factor axes in Thurstone's system must take place entirely within the common-factor space, and the common-factor axes must not leave that space and encroach upon the specifics. In Figure 18, therefore, no rotation, in Thurstone's sense, of the g axis can be made (since the common-factor space is a

line), except, indeed, reversing its direction and measuring stupidity instead of intelligence.

In Figure 19 the common-factor space is a plane, and the axes *a* and *b* can be rotated in this plane, like the hands of a clock fixed permanently at right angles to one another. When the positive directions of *a* and *b* enclose all the vector projections, as they do in our figure, then all the loadings are positive. The position shown would, therefore, fulfil this desire of Thurstone's. Moreover, one of the loadings could be made zero, by rotating *a* and *b* until *a* coincides with line 1 (when *b* will have no loading in Test 1), or until *b* coincides with line 4 (when *a* will have no loading in Test 4).

When there are three common factors, the common-factor space is an ordinary 3-space. The three common-factor axes divide this space into eight octants. Rotating them until all the loadings are positive means until all the projections of the test vectors are within the positive octant. This will always be nearly possible if the correlations are all positive. Moreover, it is clear that we can always make at any rate some loadings zero. In the common-factor 3-space we can move one of the axes until it is at right angles to two of the test projections, in which tests that factor will then have no loading. Keeping that axis fixed, we can then rotate the other two axes round it, seeking for a position where one of them is at right angles to some test. The number of zero loadings obtainable will clearly be limited unless the configuration of the test vectors happens to lend itself to many zeros. We shall see later that Thurstone seeks for teams of tests which do this.

Although Thurstone makes his rotations exclusively within the common-factor space, keeping the specifics sacrosanct at their maximum variance, there is, of course, nothing to prevent anyone who does not hold his views from rotating the common-factor axes into a wider space, and increasing the number of common-factor axes at the expense of the specific variance, until ultimately we reach as many common factors as we have tests, and no specifics.

15. *The geometrical picture of centroid analysis.*—Think of a sheaf of lines representing a number of tests, with

angles corresponding to the correlations. Centroid analysis means (if unities are used in the diagonal cells) finding a line in the middle of this sheaf—at the centroid or resultant—something like the stick in the middle of the ribs of a slightly opened umbrella, except that our test lines are not regularly spaced like those ribs.

All this is in a space of as many dimensions as there are tests, and it is not possible to make a drawing. But if the reader will be tolerant, we can make one of our "super-perspective" drawings showing a sheaf of test lines (see Figure 20) which must be imagined as being in a multi-dimensional space. The centroid line OC is the line along which the point O would move if each test line were a force—all equal—pulling O. It is exactly like the parallelogram of forces on a multi-dimensional scale. The dots on the test lines are at unit distance from O. (They have been joined by lines only in order to make the figure look more solid.) The loadings of the tests in the first centroid factor are the projections of these unit distances on to OC—this is when unities are used in the diagonal cells. The summation process gives, arithmetically, these projected distances along OC.

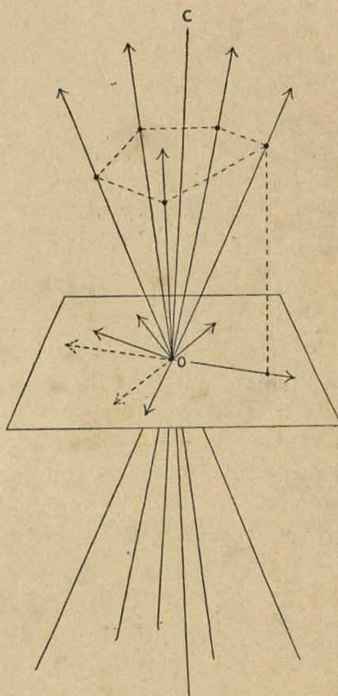


Figure 20.

The next part of the arithmetical process consisted in removing that part of the correlation coefficients explained by the first factor loadings. This means, in our space diagram, that the dimension parallel to OC is abolished, and all the test lines are projected on to a space at right angles to OC and of one dimension less than the original,

$(n - 1)$ dimensions instead of n, if n be the number of tests.

We have had perforce to draw our diagram as though it were in a three-fold instead of an n-fold space : and for this new $(n - 1)$-fold space we have drawn an ordinary plane, like a drawing-board, at right angles to *OC*, and projected the five test lines on to it. The next thing is to find the centroid of these five directed lines, these vectors, on the drawing-board. But we find at once that they are in equilibrium. If they were forces, the point *O* would not move. That is because *OC* is indeed the centroid of the original lines. This fact of equilibrium corresponds to the fact that the columns of residues add up to zero.

To get over this, in the arithmetic, we changed the signs of some rows and corresponding columns, till, if possible, all cells were positive. (These cells of the residues are the cosines of the angles on the drawing-board, some of which are clearly obtuse, with negative cosines.) This reversal of signs in the arithmetic corresponds, in our diagram, to reversing some of the vectors on the drawing-board, till they again form a sheaf, as close as possible. Two are shown as reversed in our figure, and most of the angles are now acute, most of the cosines positive. It is desirable to make the sheaf as compact as possible, corresponding to making as many cells positive as possible.

The centroid of the resulting sheaf of vectors (or forces) is the second factor. Its dimension is next abolished, by projection on to a space of $(n - 2)$ dimensions, and so on, and so on. Our possibility of following this in a drawing is beyond delineation, but if the reader will in imagination conceive of our first sheaf of test lines being in n dimensions, and being step by step projected on to spaces of $(n - 1)$, $(n - 2)$ and lesser dimensions, he will have a picture corresponding to the arithmetical summation process and the sign reversals in the residues.

For simplicity we have above supposed that unities were being left in the diagonal cells, in which case as many common factors would emerge as there were tests, and there would be no specifics. If communalities are inserted and the rank of the matrix of correlations reduced, there will be fewer *common* factors. Our diagram would then be in the common-factor space and, indeed, can still serve, if we suppose the distances from O to the dots on the test lines to be not unity, but the square roots of the communalities, and the angles to be the projections of those between the full test lines. With that change, our diagram would be one for the communal parts of five tests with three common factors, represented by OC, by the resultant of the vectors on the drawing-board (after the reversals to destroy the equilibrium), and by a third line also on the drawing-board, at right angles again.

16. *Principal components.*—The object of using centroids as axes in the above process is to obtain axes in diminishing order of importance as describers of the test lines. In the current jargon, they each " take out " as much variance as possible at each step—or rather, not quite as much as possible, though nearly so. There is another set of lines which actually do take out as much as possible. They are the lines corresponding to the axes of the ellipsoid of Figure 15, or the more general ellipsoids of higher dimensions. The centroid OC in our Figure 20 is in such a position that the sum of the squares of the vertical distances of the test dots to it is very small, nearly as small as possible. Another line, however, quite close to OC and corresponding to the major axis of the ellipsoid, makes this sum of squares an absolute minimum, and the sum of squares of the loadings of the factor a maximum.

In Section 5 above we spoke of converting the ellipsoid of our Figure 15 into a sphere by swivelling the three test lines nearer to each other till the cosines of their angles correspond to the correlation coefficients, and the test lines take up positions such as they have in our Figure 20. When this is done, the major axis of the ellipsoid takes up a position among the test lines, quite near to the centroid but not quite coinciding, and with the property of maximizing the " variance taken out." Similarly, the other principal axes of the ellipsoid, when the change is made in the space, replace for the better the later centroids of the simpler process. The arithmetical method of calculating their loadings is explained in our next chapter.

PRINCIPAL COMPONENTS

1. *A historical accident.*—By a historical accident, the method of principal components is associated in the minds of psychologists with analyses in which unities, and not communalities, are used in the diagonal cells of the square table of correlations. The centroid method can, however, equally well be used on such a table, giving the centroids of the complete test vectors in the whole test space : and the principal components of the communality vectors, in the common-factor space, can be found, using communalities in the diagonal cells, by the same iterative process as we are about to describe. As, however, this method was originally used on unit entries, we shall first make a principal components analysis of the whole tests of the example already used for the centroid process. Later we shall analyse the communality vectors by the same process (page 118).

2. *A calculation.*—The actual calculation of the loadings of principal components requires, for its complete understanding, a grasp of the method of finding algebraically the

1·0	·4	·4	·2	·8	·78	·775
·4	1·0	·7	·3	1·0	1·00	1·000
·4	·7	1·0	·3	1·0	1·00	1·000
·2	·3	·3	1·0	·7	·65	·637
·80	·32	·32	·16			
·40	1·00	·70	·30			
·40	·70	1·00	·30			
·14	·21	·21	·70			
1·74	2·23	2·23	1·46			
·780	·312	·312	·156			
·400	1·000	·700	·300			
·400	·700	1·000	·300			
·130	·195	·195	·650			
1·710	2·207	2·207	1·406			

principal axes of an ellipsoid, a problem which will be found dealt with in three dimensions in any text-book on solid geometry. We give an account of this, for n dimensions, in the Appendix. Here we shall only explain Hotelling's (1933) ingenious iterative method of doing this arithmetically, by means of an example, for which we shall use the matrix of correlations already employed in Chapter V to illustrate the centroid method (see page 108).

Hotelling's arithmetical process then begins with a guess at the proportionate loadings of the first principal component. Practically any guess will do—a bad guess will only make the arithmetic longer. We have guessed ·8, 1, 1, ·7, the numbers to be seen on the right of the matrix, because these numbers are roughly proportional to the sums of the four columns, and such numbers usually give a good first guess.

Each row of the matrix is then multiplied by the guessed number on its right, giving the matrix below the first one, beginning with ·80. We then take, as our second guess, numbers proportional to the sums of the columns of *this* matrix,* namely—

	1·74	2·23	2·23	1·46
giving	·78	1	1	·65

That is, we divide the sums of the columns by their largest member, and use the results as new multipliers. They are seen placed farther on the right of the original matrix. It is unusual for two of them to be of the same size—that is a peculiarity of our example.

It is always the original matrix whose rows are multiplied by each improved set of multipliers. The above set gives the next matrix shown, that beginning with ·780, and the sums of *its* columns—

1·710 2·207 2·207 1·406

give a third guess at the multipliers, namely—

·775 1 1 ·637

* When a calculating machine is being used, this matrix will not be actually written down—the column sums will be arrived at on the machine.

And so the reiteration goes on, and the reader, who is advised to carry it a stage farther at least, would find if he persevered that the multipliers would change less and less. If he went on long enough, he would reach this point (usually, however, far fewer decimals are sufficient):

	1·0	·4	·4	·2	·772865
	·4	1·0	·7	·3	1·000000
	·4	·7	1·0	·3	1·000000
	·2	·3	·3	1·0	·629811
	·772865	·309146	·309146	·154573	
	·400000	1·000000	·700000	·300000	
	·400000	·700000	1·000000	·300000	
	·125962	·188943	·188943	·629811	
	1·698827	2·198089	2·198089	1·384384	
giving	·772865	1	1	·629813	

that is, totals in exactly the same proportion as the multipliers. These final multipliers (or earlier ones if the experimenter is content with less exact values) are then proportionate to the loadings of the first principal component in the four tests. They have, however, to be reduced until the sum of their squares equals the largest total, 2·198089, which is called the first "latent root" of the original matrix. This is done by dividing them by the square root of the sum of their squares and multiplying them by the square root of the latent root. They then become—

·662 ·857 ·857 ·540

The next step in Hotelling's process is similar to one with which we have already become familiar in Thurstone's method. The parts of the variances and correlations due to this first component are calculated and subtracted from the original experimental matrix. These variances and correlations due to the first component are shown at the top of the opposite page.

The residual matrix is then treated in exactly the same way as the original matrix, the beginnings of the process being shown opposite. There is no need, in this process, for sign-changing. The guessed multipliers, proportional to the sums of the columns, are not so near the truth this

	·662	·857	·857	·540		
·662	·439	·567	·567	·357		
·857	·567	·734	·734	·462	Matrix due to first principal component.	
·857	·567	·734	·734	·462		
·540	·357	·462	·462	·291		
Residual matrix	·561	− ·167	− ·167	− ·157	·3	·18
	− ·167	·266	− ·034	− ·162	− ·4	− ·38
	− ·167	− ·034	·266	− ·162	− ·4	− ·38
	− ·157	− ·162	− ·162	·709	1·0	1·00
	·168	− ·050	− ·050	− ·047		
	·067	− ·106	·013	·065		
	·067	·013	− ·106	·065		
	− ·157	− ·162	− ·162	·709		
	·145	− ·305	− ·305	·792		

time, for the first one, which we have guessed at ·3, and which reduces after one operation to ·18, goes on reducing until it becomes negative, the final values of these second loadings being as shown in the appropriate column of the following table, which also gives the loadings of the third and fourth factors, obtained in the same way. The variances and correlations due to each factor in turn are subtracted from the preceding residual matrix and the new residual matrix analysed for the next factor :

Factor	*I*	*II*	*III*	*IV*	*Sum of Squares*
Test 1	·662218	− ·323324	·675967	·	1
,, 2	·856836	− ·135197	− ·312332	− ·387298	1
,, 3	·856836	− ·135197	− ·312332	·387298	1
,, 4	·539645	·826092	·162323	·	1
Sum of squares *	2·198090	·823526	·678383	·300000	4
Percentages	55·0	20·6	16·9	7·5	100

* These four quantities are, in the Hotelling process, what are called the " latent roots " of the matrix. Their product gives the value, ·3684, of the determinant of the matrix of correlation coefficients.

An alternative method of finding principal components, due to Kelley, is to deal with the variables two at a time. The pair first chosen are rotated in their plane until they are uncorrelated. Then the same is done to another pair, and so on, the new uncorrelated variables being in turn paired with others, until finally all correlations are zero. (Kelley, 1935, Chapters I and VI.) A chief advantage is that the components are obtained *pari passu*, and not successively; also, in certain circumstances where Hotelling's process converges very slowly, Kelley's is quicker. The end-results are the same.

3. *Acceleration by powering the matrix.*—In a later paper Hotelling pointed out that his process of finding the loadings of the principal components can be much expedited by analysing, not the matrix of correlations itself, but its square, or fourth, eighth, or sixteenth power, got by repeated squaring (Hotelling, 1935*b*). Squaring a symmetrical matrix is a special case of matrix multiplication (see Chapter X, Section 4, page 145): it is done by finding the "inner products" (see footnote, page 74) of each pair of rows, including each row with itself, and setting the results down in order. Applying this to the correlation matrix:

$$\begin{bmatrix} 1\cdot0 & \cdot4 & \cdot4 & \cdot2 \\ \cdot4 & 1\cdot0 & \cdot7 & \cdot3 \\ \cdot4 & \cdot7 & 1\cdot0 & \cdot3 \\ \cdot2 & \cdot3 & \cdot3 & 1\cdot0 \end{bmatrix}$$

we see that the inner product of the first row with itself is 1·36; of the first row with the second, 1·14; and so on. Setting these down in order, we get for the matrix squared:

$$\begin{bmatrix} 1\cdot36 & 1\cdot14 & 1\cdot14 & \cdot64 \\ 1\cdot14 & 1\cdot74 & 1\cdot65 & \cdot89 \\ 1\cdot14 & 1\cdot65 & 1\cdot74 & \cdot89 \\ \cdot64 & \cdot89 & \cdot89 & 1\cdot22 \end{bmatrix}$$

Exactly the same process is applied to this, beginning with guessed multipliers, as we applied to the original matrix. The multipliers, however, settle down twice as rapidly towards their final values, *which are the same here as there.* We have finally:

	1·36	1·14	1·14	·64	·772865
	1·14	1·74	1·65	·89	1·000000
	1·14	1·65	1·74	·89	1·000000
	·64	·89	·89	1·22	·629811
	1·051096	·881066	·881066	·494634	
	1·140000	1·740000	1·650000	·890000	
	1·140000	1·650000	1·740000	·890000	
	·403079	·560532	·560532	·768369	
	3·734175	4·831598	4·831598	3·043003	
Ratio	·772865	1·000000	1·000000	·629812	

The "latent root," however, or largest total, 4·831598, is the square of the former latent root, 2·198090, so that its square root must be taken before we complete finding the loadings.

In exactly the same way the squared matrix may be again squared, and again and again, before we analyse it. The more we square it, the quicker the Hotelling iteration process works. The end multipliers are always the same, but the "root" is the same power of the root we need as is the matrix of the original matrix.

A still further acceleration of the process is due to Cyril Burt, who observed that as the matrix is repeatedly squared it becomes more and more nearly hierarchical, including the diagonal cells (Burt, 1937*a*). This is due to the largest factor increasingly predominating as it is "powered," especially if the largest latent root is widely separated from the others. In consequence, the square roots of the diagonal cells become more and more nearly in the ratio of the Hotelling multipliers, and form an excellent first guess for the latter. When our matrix is squared twice again, giving the eighth power, it becomes:

108·78	140·67	140·67	88·54
140·67	182·03	182·03	114·61
140·67	182·03	182·03	114·61
88·54	114·61	114·61	72·38

and the square roots of its diagonal members are—

10·429 13·492 13·492 8·508

which are in the ratio—

·7730 1 1 ·6306

very near indeed to the Hotelling final multipliers—

·772865 1 1 ·629811

Hotelling gives a method of finding the residues, for the purpose of calculating the next factor loadings, from the "powered" matrix. But it may be so nearly perfectly hierarchical that this fails unless an enormous number of decimals have been retained, and it is in practice best to go back to the original matrix and obtain the residues from it. Their matrix can in turn be squared, and so on. Other and very powerful methods of acceleration will be found described in Aitken, 1937*b*.

4. *Properties of the loadings.*—If all the principal components are calculated accurately, and if unities were used in the diagonal cells, their loadings ought completely to exhaust the variance of each test; that is, the sum of the squares of the loadings in each row should be unity. The sum of the squares of the loadings in each column equals the "latent root" corresponding to that column, and the sum of the four latent roots is exactly equal to the number of tests. Each latent root represents the part of the whole variance of all the tests which has been "taken out" by that factor. Thus the first factor "takes out" 55 per cent., the first two factors together 75·6 per cent., of the variance of the original scores. The four factors account for all the variance.

The correlations which correspond to the loadings given in the table on page 111 are obtained by finding the

" inner product " of each pair of rows. Applying this to the table we find the correlation r_{24}, say, to be—

$$\cdot 856836 \times \cdot 539645 - \cdot 135197 \times \cdot 826092 - \cdot 312332 \times \cdot 162323 - \cdot 387298 \times \text{zero} = \cdot 300000$$

In this way the loadings of the four principal components will exactly reproduce the correlations we began with. If, however, we have stopped the analysis after we have found only two principal components (or factors), these two would have reproduced the correlations only approximately. For example, for r_{24} we should only have—

$$\cdot 856836 \times \cdot 539645 - \cdot 135197 \times \cdot 826092 = \cdot 350702 \text{ instead of } \cdot 300000$$

Before we leave the table of loadings, we may note that the signs of any column of the loadings can be reversed without changing either the variances or the correlations. Reversing the signs in a column merely means that we measure that factor from the opposite end, as we might rank people either for intelligence or stupidity and get the same order, but reversed. We will usually desire to call that direction of a factor positive which most conforms with the positive direction of the tests themselves, and therefore we will usually make the largest loading in each column positive.

All the loadings of the first principal factor are, in an ordinary set of tests, positive. Of the other loadings, about half are negative.

5. *Calculation of a man's principal components.*—Factors obtained by using unities, and not communalities, in the diagonal cells have an important advantage. *They can be calculated exactly* from a man's scores, whereas communality factors can only be estimated. This is because the former are never more numerous than the tests, whereas the latter, including the specifics, are always more numerous than the tests. For the former, therefore, we always have just the same number of equations as unknowns, whereas we have more unknowns than equations when communalities are used.

We have hitherto given the analysis of tests into factors in the form of tables of loadings. But we can alternatively write them out as "specification equations," as we shall call them. Thus the table on page 111 would be written—

$$z_1 = \cdot 662218\gamma_1 - \cdot 323324\gamma_2 + \cdot 675967\gamma_3$$
$$z_2 = \cdot 856836\gamma_1 - \cdot 135197\gamma_2 - \cdot 312332\gamma_3 - \cdot 387298\gamma_4$$
$$z_3 = \cdot 856836\gamma_1 - \cdot 135197\gamma_2 - \cdot 312332\gamma_3 + \cdot 387298\gamma_4$$
$$z_4 = \cdot 539645\gamma_1 + \cdot 826092\gamma_2 + \cdot 162323\gamma_3$$

Here z_1, z_2, z_3, and z_4 stand for the scores in the four tests, measured in standard units ; that is, measured from the mean in units of standard deviation. The factors γ_1, γ_2, γ_3, and γ_4 are also supposed to be measured in such units. These specification equations enable us to calculate any man's standard score in each test if we know his factors, and since there are just as many equations as factors, they can be solved for the γ's and enable us to calculate, conversely, any man's factors if we know his scores in the tests. The solution to these Hotelling equations for the γ's happens to be peculiarly simple, as we shall prove in the Appendix, Section 7. It is as follows—

$$\gamma_1 = (\ \cdot 662218z_1 + \cdot 856836z_2 + \cdot 856836z_3 + \cdot 539645z_4) \div 2\cdot 198090$$
$$\gamma_2 = (-\cdot 323324z_1 - \cdot 135197z_2 - \cdot 135197z_3 + \cdot 826092z_4) \div \cdot 823526$$
$$\gamma_3 = (\ \cdot 675967z_1 - \cdot 312332z_2 - \cdot 312332z_3 + \cdot 162323z_4) \div \cdot 678383$$
$$\gamma_4 = (\ - \cdot 387298z_2 + \cdot 387298z_3\ ) \div \cdot 300000$$

The table on page 111, therefore, serves a double purpose. Read horizontally it gives the composition of each test in terms of factors. Read vertically it gives the composition of each factor in terms of tests, if we divide the result by the root at the foot of the column.*

Suppose, for example, that a man or child has the following scores in the four tests—

1·29 ·36 ·72 1·03

This is evidently a person above the average in each test, since the scores are all positive. His factors will be

* If the analysis has been performed with "reliabilities" in the diagonal cells instead of units, the statement in the text still holds (Hotelling, 1933, 498). If on correlations corrected for "attenuation," the matter is more complicated (*ibid.* 499–502).

obtained by substituting these scores for the z's in the above equations, with the result—

$$\gamma_1 = 1{\cdot}062504$$
$$\gamma_2 = {\cdot}349441$$
$$\gamma_3 = 1{\cdot}034624$$
$$\gamma_4 = {\cdot}464757$$

(Of course, in practical work six decimal places would be absurd. They are given here because we are using this artificial example to illustrate theoretical points, in place of doing algebraic transformations, and they need, therefore, to be exact.)

If these values for the factors are now inserted in the specification equations opposite, the scores z in the test will be reproduced exactly (1·29, ·36, ·72, and 1·03).

Notice, too, that if we have stopped our analysis at less than the full number of principal components using unities in the diagonal cells, we can nevertheless calculate these factors for any person exactly. As soon as we have the first column of the table on page 111, we can calculate γ_1 for anyone whose scores z we know.

Had we done this with the person whose scores are given above, we should have summarized his ability in these four tests by the one statement—

$$\gamma_1 = 1{\cdot}062504$$

This would have been an incomplete statement, but it is the best single statement that can be arrived at.

6. *Principal components in the common-factor space.*—Exactly the same iterative Hotelling process for finding the principal components, the principal axes, of the ellipsoids of density of the person-points can be applied to the table of correlations with communalities in the diagonal cells. The ellipsoidal swarm of person-points, in the full test space with orthogonal axes for the tests, remains an ellipsoidal swarm (though one of fewer dimensions) when projected on to the common-factor space. The mathematical reader will know this, or can work it out. The non-mathematical reader knows it well enough in the number of dimensions he is personally acquainted with: e.g. an egg, which is an ellipsoid of three dimensions, throws a shadow on a wall

which is an ellipse, i.e. an ellipsoid of two dimensions. We shall now analyse the same set of correlation coefficients using the communalities ·26̇, ·7, ·7, ·15, which we know, from Chapter V, page 77, reduce the rank of the matrix to *two*, and give an analysis with only two common factors. We found on page 78 the two *centroid* common factors. We shall now find the two *principal components* and find them very similar.

7. *Calculation with communalities*

·2667	·4	·4	·2	·7	·59	.	.	·5913
·4	·7	·7	·3	1	1	.	.	1
·4	·7	·7	·3	1	1	.	.	1
·2	·3	·3	·15	·5	·45	.	.	·4435
1·0867	1·83	1·83	·815					

Taking ·7, 1, 1, ·5 as a first guess at the multipliers, we find the weighted sums of the columns to be as shown, and on dividing through by 1·83 we get the next set of multipliers ·59, 1, 1, ·45. Continuing in this way, we arrive quite soon at ·5913, 1, 1, ·4435, which, when used as weights, reproduce themselves. When reduced until the sum of their squares equals 1·7696 (the largest column total with these weights), the loadings are—

·4929 ·8336 ·8336 ·3697

Subtracting the cross-products of these from the original matrix, and operating on the residues in exactly the same iterative way, we get for the second factor loadings—

·1540 — ·0712 — ·0712 ·1153, and no residues.*

If we compare these principal component loadings with the centroid loadings (page 78) obtained with the same communalities, we see that they are very similar. But the sum of squares of the loadings of the first principal component (1·7694) is slightly larger than the same sum for the first centroid loadings (1·7652). The principal compon-

* The sums of squares of the loadings (1·7694 and ·0471) are the two first latent roots of the matrix with communalities. The other two latent roots are zero. The sum of the latent roots equals the sum of the communalities, the " trace " of the matrix as it is called.

ents take out at each stage the maximum possible variance (sum of squares of loadings). The centroids nearly do so if the sign-changing is carefully done, but not quite. The centroids can best be looked at as approximations to the principal components, more easily calculated. In a battery of many tests, say two dozen, and with any given communalities, the principal component process (" weighted summation ") will take out more variance in, say, six factors, and leave smaller residues, than will centroid factors.* But with the kind of data available in psychology, this advantage does not outweigh the disadvantage of longer calculation.

8. *Iterative methods.*—Both in the above Hotelling calculation, and in our discussion of communalities on page 88, we have seen examples of iterative processes, where a first guess at certain constants gives results which can be used as a better guess, which gives results which can be used as a still better guess, which gives . . . and so on and so on, until the stage is reached where the same constants emerge as were put in. This sort of process, where repetition after repetition converges to a steady result giving some maximum or minimum value to some quantity, is not uncommon in mathematics and is rather mysterious and magical to the layman. An analogy will perhaps assist understanding. Robinson Crusoe wants to make a lathe, but he has no wheels and spindles, and to make wheels and spindles he needs a lathe! He can, however, whittle crude makeshift wooden wheels, etc., with a knife, and make a crude lathe with them, with which lathe he can make somewhat better wheels and therefore a somewhat better lathe, with which he can make still better wheels . . . and so on, till he reaches perfection.

* If the Hotelling process is used with guessed communalities, and the whole is iterated (as was done with centroids on page 88) the communalities will converge to a set minimizing the sum of squares of the residuals for a given number of factors. The maximum likelihood method of Chapter IX arrives at communalities (I understand from Dr. Lawley) which minimize a *weighted* sum of squares of the residuals, each weight being the product of the reciprocals of the two specific variances concerned.

CHAPTER VIII

TESTING RESIDUES FOR SIGNIFICANCE

1. *The object of factorial analysis.*—As was said in the first section of Chapter I, the objects of factorial analysis are both practical and theoretical. The practical desire is to reduce the description of a man's mind* to a comparatively few quantitative statements, instead of an unwieldy record of innumerable test scores, with a view to giving vocational or educational advice. The hope, on the theoretical side, is that the "factors" found may form the structure of a theory of mind: and there are some who hope that physiological or neurological bases may be found for them. Our concern in this chapter is with the first point: how to reduce the number of "factors" without sacrificing any significant fraction of the information. The insertion of communalities in the diagonal cells of a table of correlations is by many looked upon as one way of doing this, since it reduces the number of *common* factors. Simultaneously, however, it creates and maximizes the influence ascribed to specific factors, and the total number of factors is increased, not diminished. This will not be discussed in the present chapter, which is concerned with another way of reducing the number of factors, applicable whether communalities or full variances are analysed. If the idea of communalities and specifics had never occurred to anyone, it would still have been possible to reduce the number of *significant* common factors to a number less than the number of tests. Each principal component, found as described in Chapter VII, causes the remaining residues to be as small as can be: and the centroid factors of Chapter V are nearly as good, if the sign-changing is done properly. If, after a few such factors have been extracted, the residues are so small as to be statistically negligible, we

* Or of other objects of study, say in agriculture or in engineering. See Chapter XII, Section 7.

might as well stop the analysis, content with the few factors extracted. We need, therefore, some test of statistical significance, applicable to such residual correlations, to know if they are negligible.

2. *The general idea of significance.*—The general principle of such a test of significance is this, that if the residues we have found, or in practice some function of them, could only rarely have been produced by the action of chance sampling, we will assume that they are not due to sampling but to another factor. How we define " rarely " depends on circumstances. Usually in psychology " once in twenty times " (the 5 per cent. point as it is called) is rare enough to justify taking out another factor. The principle is straightforward enough, the mathematical difficulty of finding formulæ for calculating the chances, however, very great, even for principal components with full variances, and insuperable when the centroid method is used with guessed communalities. In consequence, a number of rule-of-thumb criteria have been put forward, to decide when to stop factorizing.

3. *Empirical rules for the number of factors.*—Thurstone (1938*a*, 65 *et seq.*) discusses some of the earlier ones. A criterion which appeals to common sense is based simply on the algebraic sum of the residuals (excluding the diagonal cells) after as many as possible of their signs have been made positive by the process described in Chapter V (page 71). As long as this sum goes on sinking, factorization is continued. When it flattens, the last factor taken out is rejected and the process stopped. Mosier (1939) found this the best of five plans he tried, though none was wholly satisfactory.

Ledyard Tucker's criterion is that the ratio of the sums of the absolute values of the residuals, including the diagonal used, just after and just before the extraction of a factor must be less than $(n - 1)/(n + 1)$ where n is the number of tests.

Coombs' criterion depends upon the number of negative signs left among the residuals after everything has been done to reduce them by sign-changing, in the centroid process. If they are few, another factor may be extracted.

More exactly, the permissible number is given in this table :

Number of tests . .	10	15	20	25	30
Negative signs . .	31	79	149	242	358
Standard error . .	5	7	10	12	15

A fuller table is given in Coombs' article (1941).

An example of the use of these two will be found in Blakey (1940, 126).

Quinn McNemar (1942), who considers both of the above inadequate, gives a formula which includes N the size of the sample. He takes out factors until σ_1 reaches or falls below $1/\sqrt{N}$, where

$$\sigma_1 = \sigma_s \div (1 - M_{h^2}),$$
$$\sigma_s = \text{st. dev. of the residuals after } s \text{ factors},$$
$$M_{h^2} = \text{mean communality for } s \text{ factors}.$$

Others go on until the distribution of the residuals ceases to be significantly skew (Swineford, 1941, 378). Reyburn and Taylor (1939) divide the residuals by the probable errors of the original coefficients, and plot a distribution of the results disregarding signs. If it is significantly different from a normal curve of the same area and with standard deviation 1·4825, they take out more factors. Swineford (1941, 377) finds the correlation between the original correlations and the corresponding residuals and takes out factors till it is not significant.

Another method is based on the sinking of the factor loadings with each successive factor instead of on the dying away of the residuals. Guilford and Lacey (1947 in a U.S. Air Force report) stop factorizing when the product of the two highest factor-loadings falls below $1/\sqrt{N}$.

P. E. Vernon, in a privately circulated manuscript, has tested some two dozen methods, as applied when the centroid or simple summation method of analysis is used with communalities, on two analyses of actual data, on 645 and 994 cases respectively (Vernon, 1947). His final advice is to use the methods of Guilford and Lacey (product of the two highest factor loadings) and of Mosier (sum of the residuals), together with Burt's empirical formula for the standard error of each factor loading—

$$\frac{(1 - l^2)\sqrt{n}}{\sqrt{N(n - s + 1)}}$$

where l = loading, N = number of persons, n = number of tests, s = the ordinal number of the factor. If half the loadings of a factor fall below twice their standard errors thus found, Vernon recommends rejection of the factor.

If these three methods do not agree, Vernon would proceed to calculate McNemar's σ_1 (opposite), and would decide on the evidence of the four criteria, taking out another factor if doubtful.

4. *More exact methods.*—The earliest method was to compare each residue with the standard error of the original correlation coefficient and cease factorizing when the residues all sank below twice these standard errors. But the use of the formula for the standard error of r is now frowned upon because of the skewness of the distribution.

Moreover, sampling errors in the correlation coefficients, being themselves correlated, produce further factors ; and the above-mentioned test tended to stop the analysis too soon (Wilson and Worcester, 1939). These further factors must be taken out in order to give elbow room for rotation of the axes to some psychologically significant position. For the error factors are not concentrated in the last factors taken out, but have been entangled with all. Usually more factors have to be taken out than can be expected, on rotation, to yield meaningful psychological factors, but all the dimensions are required nevertheless for the rotations. In geometrical terms, some of the dimensions of the common factor space will be due to sampling error, but not the particular dimensions indicated by the directions of the last factors to be extracted. In terms of Hotelling's plan, the whole ellipsoid is distorted ; its small major axes are not necessarily due entirely to sampling, nor its large ones free from it. A χ^2 method is described by Wilson and Worcester (1939, 139) which is, however, laborious when the number of tests is large. See also Burt (1940, 338–40). Lawley (1940, 76 *et seq.*) repeated Wilson and Worcester's criticism and developed an accurate criterion described in the next chapter. This is probably the best plan to use in any research where great accuracy is necessary. And it is for the case where communalities are employed. It is, however, only legitimate when the factor loadings have been found by Lawley's application of the method of maximum likelihood.

Principal components lend themselves to exact treatment when full unities are used, i.e. there are no specifics assumed. Hotelling himself (1933, 437–41) discusses the matter of the number which are significant. Davis (1945) shows how to find the reliability of each principal component from the reliabilities of the tests, and finds that it may happen that a later component is more reliable than an earlier one.

5. *M. S. Bartlett's test of significance for principal components.*—Recently (Bartlett, 1950) a method has been described for deciding the significance of principal component factors which, while it is unlikely, in its present form at least, to be usable in any ordinary cases, ought to be briefly described here. It is highly desirable that exact methods, or methods where the assumptions made and the approximations permitted are clearly realized and set out, should gradually replace those based on experience only. Bartlett's method depends upon the latent roots of the matrix of correlation coefficients with unity in each diagonal cell—it is not applicable to communalities.

Latent roots have been mentioned on page 111, where they appear as the sums of squares of the loadings of the tests in each principal component. In the example there used, their values are—

$$\begin{aligned} \lambda_1 &= 2{\cdot}198 \\ \lambda_2 &= \ {\cdot}824 \\ \lambda_3 &= \ {\cdot}678 \\ \lambda_4 &= \ {\cdot}300 \end{aligned}$$

They are equal in number to the tests, and their sum also is exactly 4. Bartlett forms quantities R_i as follows :

		$\log_e R$
$R_1 = \lambda_4 \times \dfrac{1}{\lambda_4}$	$= 1$	
$R_2 = \lambda_3\lambda_4\left(\dfrac{2}{\lambda_3 + \lambda_4}\right)^2$	$= \cdot 8506$	$- 0\cdot 16182$
$R_3 = \lambda_2\lambda_3\lambda_4\left(\dfrac{3}{\lambda_2 + \lambda_3 + \lambda_4}\right)^3$	$= \cdot 7734$	$- 0\cdot 25696$
$R_4 = \lambda_1\lambda_2\lambda_3\lambda_4$	$= \cdot 3684$	$- 0\cdot 99858$

and of these we require the *natural* logarithms, which are 2·3026 times the usual logarithms to the base ten. They are given above. These logarithms, multiplied by a certain coefficient, are an approximation to χ^2 for the successive factors. The coefficient is—

$$- n + \frac{2p + 5}{6} + \frac{2k}{3}$$

where n is the number of persons tested less one, p is the number of latent roots, i.e. of tests, and k is the number of factors already dealt with, i.e. it takes in turn the values 0, 1, 2, . . .

In our example $p = 4$. If we assume that the number of persons tested was 20, so that Bartlett's $n = 19$, we can make this table :

k	*d.f.*	χ^2	5 *per cent. level*
0	3 + 2 + 1	$- 16\cdot 833 \times (- \cdot 99858) = 16\cdot 8095$	12·59
1	2 + 1	$- 16\cdot 167 \times (- \cdot 25696) = 4\cdot 1542$	7·82
2	1	$- 15\cdot 500 \times (- \cdot 16182) = 2\cdot 5082$	3·84

The quantities in the last column are to be obtained from a χ^2 table, entered with the number of degrees of freedom (*d.f.*) shown. Only the first factor is significant (16·8095 being greater than 12·59).

If we had assumed 29 children ($n = 28$) we should have been puzzled by a peculiar result. The three values of χ^2 are then 25·80, 6·47, and 3·96, so that it looks as though the first factor and the third factor are significant, with the factor in between not significant!* But Bartlett warns (1950, 78) that this χ^2 test is only valid if the roots already removed are significant. As soon as we come to a non-significant factor, the later factors are also non-significant. The last factor of all is not dealt with. "Merely the correlation structure of the variables is being investigated in its relation to variance," says Bartlett (page 80). "For this reason no significance can ever be attached to the last root, for it would be equivalent to asking for the correlation structure of a single variable."†

* Compare the report by Davis (1945) that a later component may be more reliable than an earlier one.

† In a later paper (*B.J.P. Statist.* **4**, p. 1) Bartlett warns that after one or more significant components have been eliminated it is safer to take as the number of degrees of freedom

$$\tfrac{1}{2}\ (p-k-1)\ (p-k+2)$$

instead of

$$\tfrac{1}{2}\ (p-k)\ (p-k-1)$$

as used above. This would increase the degrees of freedom in the second line of the analysis on page 125 from 3 to 5, and in the third line from 1 to 2, and raise the 5 per cent. level.

CHAPTER IX

THE MAXIMUM LIKELIHOOD METHOD OF ESTIMATING FACTOR LOADINGS *

(by D. N. Lawley)

1. *Basis of statistical estimation.*—In recent times attempts have been made to introduce into factorial analysis statistical methods developed in other fields of research. In particular the method of statistical estimation put forward by Fisher (1921, page 323 *et seq.*), and termed the method of maximum likelihood, has been applied by Lawley (1940, 1941, 1943) to the problem of estimating factor loadings. This method has the property of using the largest amount of available information contained in the data and gives "efficient" estimates, where such exist, of all unknown parameters, i.e. estimates which, roughly speaking, are on the average nearer the true values than those obtained by other, "inefficient," methods of estimation.

Before using the maximum likelihood method for estimating factor loadings it is necessary to make certain initial assumptions. We assume that both the test scores and the factors, of which they are linear functions, are normally distributed throughout the population of individuals to be tested. This assumption of normality has been the subject of some criticism, but in practice it would appear that departure from strict normality of distribution is not very serious. It is also necessary to make some hypothesis concerning the number of general factors which are present in addition to specifics. We shall later on show how this hypothesis may be tested, and how it may be determined whether the number assumed is, in fact, sufficient to account for the data.

2. *A numerical example.*—In order to illustrate the calcu-

* For a detailed exposition of the arithmetical procedure of Lawley's method, with checks, see Emmett (1949).

lations needed we shall reproduce an example used by Lawley (1943*b*), where eight tests were given to 443 individuals. The table below gives the correlations between the eight tests, unities having been placed in the diagonal cells. In this example the hypothesis made is that two general factors, together with specifics, are sufficient to account for the observed correlations.

	1	2	3	4	5	6	7	8
1	1·000	·312	·405	·457	·500	·350	·521	·564
2	·312	1·000	·460	·316	·279	·173	·339	·288
3	·405	·460	1·000	·394	·380	·258	·433	·323
4	·457	·316	·394	1·000	·460	·222	·516	·486
5	·500	·279	·380	·460	1·000	·239	·441	·417
6	·350	·173	·258	·222	·239	1·000	·302	·262
7	·521	·339	·433	·516	·441	·302	1·000	·547
8	·564	·288	·323	·486	·417	·262	·547	1·000

The method of estimation about to be described is one of successive approximations. Each successive step in the calculations gives a set of factor loadings which are nearer to the final values than those of the previous set. To start the process it is only necessary to guess or to find by some means (e.g. by a centroid analysis) first approximations to the factor loadings. Any set of figures within reason will serve the purpose, though, of course, the better the approximation the fewer steps in the calculation will be needed. For illustration we shall take as first approximations to the factor loadings the set of values given below :

Trial	Tests							
loading in	1	2	3	4	5	6	7	8
Factor I	·73	·50	·66	·66	·62	·40	·73	·70
Factor II	·17	−·27	−·47	·08	·06	·02	·10	·29
Specific variance	·4382	·6771	·3435	·5580	·6120	·8396	·4571	·4259

Under the loadings are written the corresponding first approximations to the specific variances (the total variance of each test being taken to be unity). They are as usual found by subtracting from unity the sums of squares of the loadings for each test.

The calculations necessary for obtaining second approximations to the loadings in factor I may now be set out as follows :

(*a*)	1·666	·738	1·921	1·183	1·013	·476	1·597	1·644
(*b*)	5·647	3·895	5·132	5·129	4·830	3·100	5·647	5·412
(*c*)	4·917	3·395	4·472	4·469	4·210	2·700	4·917	4·712
		$h_1^2 = 45{\cdot}724$			$1/h_1 = 0{\cdot}14789$			
(*d*)	·727	·502	·661	·661	·623	·399	·727	·697

The first row of figures, row (*a*), is found by dividing the trial loadings in factor I by the corresponding specific variances. The figures in row (*b*) are then given by the inner products (see footnote, page 74) of row (*a*) with the successive rows (or columns) of the correlation table printed above, and row (*c*) is obtained by subtracting from the figures in row (*b*) the corresponding loadings in factor I. The quantity h_1^2 is given by the inner product of rows (*a*) and (*c*), and hence, taking the square root of the reciprocal of this quantity, we find $1/h_1$. Finally, row (*d*) is obtained by multiplying the figures in row (*c*) by $1/h_1$, or ·14789. The resulting numbers are then second approximations to the loadings of the tests in factor I.

The most direct way of obtaining second approximations to the loadings in factor II is to find the residual matrix which results from removing the effect of factor I, and to treat it in the same way as the original matrix, using this time the trial loadings in factor II. A less direct but considerably shorter method may, however, be obtained by using once more the original matrix and modifying the process slightly. The necessary calculations are as shown below :

(*e*)	·388	−·399	−1·368	·143	·098	·024	·219	·681
(*f*)	·330	−·560	−·980	·150	·113	·038	·190	·580
		$p_1 = -{\cdot}0234$						
(*g*)	·177	−·278	−·495	·085	·068	·027	·107	·306
		$k_1^2 = 1{\cdot}1080$			$1/k_1 = {\cdot}9500$			
(*h*)	·168	−·264	−·470	·081	·065	·026	·102	·291

Row (*e*) is found by dividing the trial loadings in factor II by the corresponding specific variances (thus, ·388 is ·17/·4382), while the numbers in row (*f*) are given by the inner products of row (*e*) with the rows of the correlation table.

The step by which row (*g*) is obtained from row (*f*) is a little more complicated than the corresponding step in the calculations for the first-factor loadings. From each number in row (*f*) we subtract not only the corresponding trial loading in factor II, but also a correction which eliminates the effect of factor I; this correction consists of the corresponding number in row (*d*) multiplied by — ·0234, the inner product of rows (*e*) and (*d*). Thus, for example, the number ·177 in row (*g*) is equal to

$$\cdot 330 - \cdot 170 - \cdot 727 \times (- \cdot 0234)$$

In general, where more than two factors are assumed to be present and where further approximations are being calculated for the loadings in the *r*th factor, there will be $(r - 1)$ such corrections to be subtracted, one for each of the preceding factors.

Having found row (*g*) the quantity k_1^2 is now given by the inner product of rows (*e*) and (*g*), from which, taking the square root of the reciprocal, we derive $1/k_1$. Row (*h*) is then obtained by multiplying the figures in row (*g*) by $1/k_1$, or ·9500. We have thus found second approximations to the loadings in factor II.

The whole cycle of calculations may now be repeated over and over again until the required degree of accuracy is reached. In practice, provided that the initial trial loadings are not too far out, one repetition of the process will usually be found sufficient. In our example the final estimates (with possible slight errors in the last decimal place) were as follows :

	Tests							
Loading in	1	2	3	4	5	6	7	8
Factor I	·725	·503	·664	·661	·623	·399	·726	·694
Factor II	·172	—·261	—·468	·087	·069	·027	·106	·291
Specific variance	·445	·679	·340	·556	·607	·840	·462	·434

Having obtained these figures, there is, of course, no objection to rotating the factors as desired in order to reach a psychologically acceptable position.

3. *Testing significance.*—A difficulty in most systems of factorial analysis is to know how many factors it is worthwhile to "take out," and to decide how many of them may be considered significant. From a statistical point of view objections can be raised against the majority of methods at present in use for this purpose. When, however, the number of individuals tested is fairly large, the maximum likelihood method provides a satisfactory means of testing whether the factors fitted can be considered sufficient to account for the data.

To illustrate this let us return to the example of the previous section. It is first of all necessary to calculate the matrix of residuals obtained when the effect of both factors is removed from the original correlation matrix. For this purpose we use the final estimates of the loadings as already given. The residual matrix, with the specific variances inserted in the diagonal cells, is as follows :

	1	2	3	4	5	6	7	8
1	(·445)	−·008	·004	−·037	·036	·056	−·024	·011
2	−·008	(·679)	·004	·006	−·016	−·021	·001	·015
3	·004	·004	(·340)	−·004	−·001	·006	·001	−·002
4	−·037	·006	−·004	(·556)	·042	−·044	·027	·002
5	·036	−·016	−·001	·042	(·607)	−·011	−·019	−·035
6	·056	−·021	·006	−·044	−·011	(·840)	·009	−·023
7	−·024	·001	·001	·027	−·019	·009	(·462)	·012
8	·011	·015	−·002	·002	−·035	−·023	·012	(·434)

We are now able to calculate a criterion, which we shall denote by w, for deciding whether the hypothesis that only two general factors are present should be accepted or rejected. Each of the above residuals is squared and divided by the product of the numbers in the corresponding diagonal cells. Thus, for example, the residual for Tests 4 and 7 is squared and divided by the product of the fourth and seventh diagonal elements, giving the result

$$\frac{(\cdot 027)^2}{\cdot 556 \times \cdot 462} = \cdot 002838$$

There are altogether 28 such terms, one for each residual, and w is obtained by forming the sum of these terms and multiplying it by 443, the number in the sample. The result is found to be 20·1.

When the number in the sample is fairly large w is distributed approximately as χ^2 with degrees of freedom given by

$$\tfrac{1}{2}\{(n-m)^2 - n - m\}$$

where n is the number of tests and m is the assumed number of factors. To test whether the above value of w is significant we now use a χ^2 table such as is given by Fisher and Yates (1938, page 27). In our case, putting $n = 8$ and $m = 2$, the number of degrees of freedom is 13. Entering the χ^2 table with 13 degrees of freedom, we find that the 1 per cent. significance level is 27·7. This means that if our hypothesis that only two general factors are present is correct, then the chance of getting a value of w greater than 27·7 is only 1 in 100. If, therefore, we had obtained a value of w greater than 27·7 we should have been justified in rejecting the above hypothesis and in assuming the existence of more than two general factors. In our case, however, the value of w is only 20·1, well below the 1 per cent. significance level. We have thus no grounds for rejection, and although we cannot state that only two general factors are present, we have no reason to assume the existence of more than two.

It must be emphasized that the method described above is not applicable if other, inefficient, estimates of the loadings are substituted for the maximum likelihood estimates. For the value of χ^2 would in that case be greatly exaggerated, causing us to over-estimate its significance. For this reason we cannot, for example, use the method for testing the significance of the residuals left when factors have been fitted by the centroid method.

4. *The standard errors of individual residuals.*—A method has now* been developed for finding the standard errors of individual residuals. This should be useful when a few of the residuals are very large, while the rest are small. In such a case one or more of the residuals may be highly

* Lawley in the *Proc. Roy. Soc., Edin.*, 1949.

significant, when tested individually, even though the value of χ^2 does not attain significance. The method ignores errors of estimation of the specific variances, which are not, however, likely to be very large provided that the number of tests in the battery is not too small.

Let us denote by l_i, m_i the estimated loadings of the i^{th} test in the first and second factors respectively (assuming the existence of only two factors). Let v_i be the specific variance of the i^{th} test, and let us write—

$$h = \sum_i \frac{l_i^2}{v_i},$$

$$k = \sum_i \frac{m_i^2}{v_i}$$

Then the standard error of the residual for the i^{th} and j^{th} tests $(i \neq j)$ is given by—

$$\sqrt{\frac{1}{N}\left(e_{ii}e_{jj} - e_{ij}^2\right)}$$

where $$e_{ii} = v_i - \frac{l_i^2}{h} - \frac{m_i^2}{k}$$

and $$e_{ij} = -\frac{l_i l_j}{h} - \frac{m_i m_j}{k}$$

This formula may, of course, be easily extended to take into account any number of factors.

Let us illustrate the use of the above formula with the same numerical example as before. If we wish to test the significance of the residual for the first and fourth tests after removing two factors, we have—

$$l_1 = \cdot 725 \qquad m_1 = \cdot 172 \qquad v_1 = \cdot 44479$$
$$l_4 = \cdot 661 \qquad m_4 = \cdot 087 \qquad v_4 = \cdot 55551$$
$$h = 6{\cdot}7185 \qquad k = 1{\cdot}0528$$

Hence $e_{11} = \cdot 33845 \quad e_{44} = \cdot 48329 \quad e_{14} = -\cdot 08554$

and $$\sqrt{\frac{1}{443}\left(e_{11}e_{44} - e^2_{14}\right)} = \cdot 0196$$

Thus the residual in question has a value of ·037 with a standard error of ·020. It is clearly not significant.

5. *The standard errors of factor loadings.*—When maximum likelihood estimation has been used, we are able to find the standard errors of not only the residuals but also the estimated factor loadings. Using the same notation as in the preceding section, the sampling variance of l_i, the loading of the i^{th} test in the first factor is (assuming the test to be standardized)—

$$\frac{1}{N}\left(1+\frac{1}{h}\right)\left\{1-\tfrac{1}{2}\left(1+\frac{1}{h}\right)l_i^2\right\}$$

and the standard error is the square root of this.

The covariance between any two first factor loadings l_i and l_j is given by—

$$\frac{1}{N}\left(1+\frac{1}{h}\right)\left\{r_{ij}-\tfrac{1}{2}\left(1+\frac{1}{h}\right)l_il_j\right\}$$

The formulæ for the variances and covariances of the subsequent factor loadings are more complex. Thus the variance of m_i, the loading of the i^{th} test in the second factor, is—

$$\frac{1}{N}\left(1+\frac{1}{k}\right)\left\{1-\left(1+\frac{1}{h}\right)l_i^2-\tfrac{1}{2}\left(1+\frac{1}{k}\right)m_i^2\right\}$$

while the covariance between m_i and m_j is

$$\frac{1}{N}\left(1+\frac{1}{k}\right)\left\{r_{ij}-\left(1+\frac{1}{h}\right)l_il_j-\tfrac{1}{2}\left(1+\frac{1}{k}\right)m_im_j\right\}$$

The results for the general case, where more than two factors have been assumed present, may be written down without difficulty. Each factor will give rise to one more term within the curly brackets than the preceding factor. It should be noted that the last of such terms, and that alone, is multiplied by $\frac{1}{2}$.

The variances and covariances of loadings in any factor are those for given values of the loadings in all preceding factors.

It must be stressed that all the above results are applicable only to the unrotated loadings.

In our numerical example, we find—

$$1 + \frac{1}{h} = 1{\cdot}14884$$

$$1 + \frac{1}{k} = 1{\cdot}9498$$

Hence the variance of l_1, for example, is

$$\frac{1{\cdot}14884}{443}\left\{1 - \tfrac{1}{2} \times 1{\cdot}14884 \times {\cdot}725^2\right\} = {\cdot}001810$$

while that of m_1 is—

$$\frac{1{\cdot}9498}{443}\left\{1 - 1{\cdot}14884 \times {\cdot}725^2 - \tfrac{1}{2} \times 1{\cdot}9498 \times {\cdot}172^2\right\} = {\cdot}001617$$

Thus the loading of test 1 in the first factor is ·725, with a standard error of—

$$\sqrt{{\cdot}001810} = {\cdot}043$$

and its loading in the second factor is ·172, with a standard error of—

$$\sqrt{{\cdot}001617} = {\cdot}040$$

6. *Advantages and disadvantages.*—To sum up: the chief advantage of the maximum likelihood method of estimating factor loadings is that it does lead to efficient estimates and does provide a means of deciding how many factors may be considered necessary. It unfortunately takes, however, much longer to perform than a centroid analysis, particularly when the battery of tests is a large one and when several factors are to be fitted. The chief labour of the process lies in the calculation of the various inner products; although in this respect it does not differ greatly from Hotelling's method of finding "principal components." The maximum likelihood method is thus likely to be most useful in cases where accurate estimation is desirable and where it is proposed to make a test of significance.

The method also possesses the advantage of being independent of the units in which the test scores are measured. The same system of factors is therefore obtained whether the correlation or the covariance matrix is analysed. The loadings in the one case are directly proportional to those in the other.

PART III

THE ROTATION OF FACTORS

CHAPTER X

THE ROTATION OF FACTORS

1. *Is rotation necessary?*—The factors or axes arrived at by the centroid process (or as principal components) are not at all the same sort of things as the Spearman system and its extensions gave. The Spearman factors, though mathematical devices are used in calculating their loadings, have psychological meaning from the first. Their names indicate this—general intelligence, the verbal factor, etc. There is no need for rotating them.

With the other kind of factor, the case is different. As first obtained, they make no claim to have psychological meaning. Their virtue is a purely mathematical virtue—they each explain, in turn, as much as possible of the variance of the tests, and arrive with as few common factors as possible at negligible residues. The loadings of the first centroid* factor are usually all positive, and it runs as a positive factor through all the tests. But it is not as a rule identical with Spearman's *g*. The succeeding centroid factors have each negative loadings in about half the tests, and are often referred to as bipolar factors. They may be looked upon as repeatedly classifying the tests into subgroups, and this classification may be expressed by a kind of family tree :

Factor I	All loadings positive			
Factor II	Positive loadings		Negative loadings	
Factor III	Positive loadings	Negative loadings	Positive loadings	Positive loadings

Not infrequently the sub-families into which this bipolar classification analyses the tests will have something psycho-

* This is the most convenient name, to avoid verbosity. But unless it is otherwise stated, may it be understood that principal components are equally referred to.

logical in common, and to that extent these factors in such cases may claim to have psychological meaning. Much depends on how the battery of tests is made up. And such bipolar classification is more natural in tests of temperament and character, where common speech has many bipolar phrases (as brave-cowardly, modest-cheeky, etc.), than in tests of an intellectual nature, though there too bipolar pairs of words are found, like clever-stupid.

Many psychologists, however, especially if they tend to look upon factors as real mental entities, even perhaps with physiological causes, find it difficult to admit all those negative loadings. A mental ability or factor, they argue, is on the whole something which helps us to do things, not hinders. A few negative loadings they can understand: but not so many as half the loadings. So they wish to turn the centroid axes into positions where most of the loadings will be positive, and moreover positions to which they can give psychological meaning, and which will be found and be recognizable in different batteries of tests. For this purpose the factor-analyst must be instructed in methods of rotating the centroid factors into new positions.

2. *Methods of rotation.*—One method, Alexander's, has already been described earlier in this book on pages 79 to 80. It was used by Alexander himself with excellent effect (Alexander, 1935), but involves assuming (*a*) that the communality of a certain test is entirely due to one factor; (*b*) that the communality of a second test is entirely due to this factor and one other; (*c*) and so on for $r - 1$ tests, where r is the number of factors. The criterion of success with this method is to see whether, when these assumptions are made, negative loadings disappear; and whether the consequent loadings of those tests about which no assumptions are made are compatible with the psychologist's *psychological* analysis of them. Alexander's assumptions, however, cannot generally be made in a usual battery of tests, and other methods of rotation are required. The simplest plan is to rotate the factors two at a time in their own plane. An example will best explain this.

3. *Two-by-two rotation.*—Let us suppose that we have the following set of loadings in eight tests for three factors:

	I_0	II_0	III_0	h^2
1	·4	·4	·1	·33
2	·7	·3	— ·4	·74
3	·7	— ·2	— ·3	·62
4	·9	— ·1	·3	·91
5	·5	·2	— ·2	·33
6	·8	— ·4	·1	·81
7	·6	·5	— ·2	·65
8	·5	— ·3	·4	·50

Suppose further that we want to rotate to positions of the three axes where there will be no negative loadings, or at least only few, and those small. We shall do this taking the axes two by two, and rotating each pair in its own plane. Take first axes I_0 and II_0, where the subscripts indicate that no rotation has yet taken place. Draw a diagram, using the loadings on I_0 and II_0 as co-ordinate axes (Figure 21). We can see at once that if we rotate the axes to new positions I_1 and II_1 they will enclose all the test points in their positive quadrant, and all the loadings on these two axes will be positive. The position is, however, not unique, for we could have rotated a little farther, or a little less, than θ and still enclosed all the points. I have taken θ as 37°, with sine θ = ·6 and cosine θ = ·8.

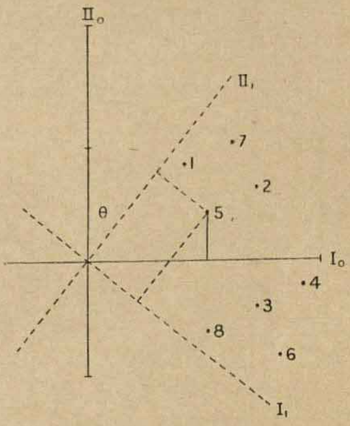


Figure 21.

Consider now the point 5. Its co-ordinates on the former axes were ·5 and ·2, and clearly its new co-ordinates are—

$$·5 \cos \theta - ·2 \sin \theta = ·28$$

and

$$·5 \sin \theta + ·2 \cos \theta = ·46$$

These can be checked approximately on the diagram, and this should always be done, at least by eye if not by measurement. The new loadings of each of the tests can be calculated in the same way, giving—

	I_1	II_1	*Sum of squares*
1	·08	·56	·32
2	·38	·66	·58
3	·68	·26	·53
4	·78	·46	·82
5	·28	·46	·29
6	·88	·16	·80
7	·18	·76	·61
8	·58	·06	·34

At this point two checks should be made: (1) The sum of the squares of the loadings of any test in these two factors should not have altered. Thus $\cdot 08^2 + \cdot 56^2$ is the same as $\cdot 4^2 + \cdot 4^2$ for the first test. (2) The inner product of any pair of rows should not have altered. Thus, for the first two tests:

$$\cdot 4 \times \cdot 7 + \cdot 4 \times \cdot 3 = \cdot 40$$

and $$\cdot 08 \times \cdot 38 + \cdot 56 \times \cdot 66 = \cdot 4000$$

It is sufficient to check only adjacent rows.

Our three axes are now I_1, II_1, and III_0, and III_0 still has negative loadings. We must therefore rotate it with one of the others, which will have its loadings further changed. Let us choose I_1 and III_0, and with their loadings make this diagram (Figure 22).

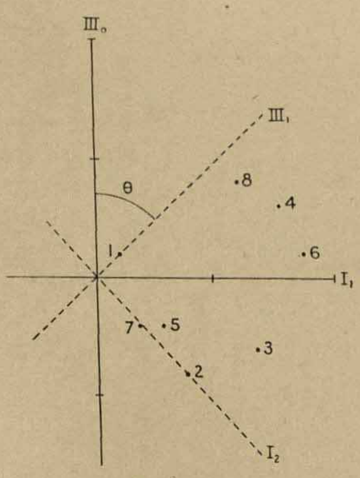


Figure 22.

A little trial with a square corner of a piece of paper shows us that we cannot rotate the axes to a position which will completely enclose all the points, though we very nearly can. We finally decide to make I_2 go exactly through point 2, whose co-ordinates in this diagram are ·38 and − ·4. The sine and cosine of θ are therefore :

$$\frac{\cdot 4}{\sqrt{\cdot 38^2 + \cdot 4^2}} \quad \text{and} \quad \frac{\cdot 38}{\sqrt{\cdot 38^2 + \cdot 4^2}}$$

or

$$\cdot 725 \text{ and } \cdot 689$$

$$(\text{check that } \cdot 725^2 + \cdot 689^2 = \text{unity})$$

The loadings of the point 5, for example, are then :

$$\cdot 689 \times \cdot 28 - \cdot 725 \times (-\cdot 2) = \cdot 338 \text{ on } I_2$$

and

$$\cdot 725 \times \cdot 28 + \cdot 689 \times (-\cdot 2) = \cdot 065 \text{ on } III_1$$

as can be approximately checked by a look at the diagram.

In the same way the other loadings on I_2 and III_1 can be found, giving the complete table :

	I_2	II_1	III_1	h^2
1	− ·017	·560	·127	·3300
2	·552	·660	·000	·7403
3	·686	·260	·286	·6200
4	·320	·460	·772	·9100
5	·338	·460	·065	·3301
6	·534	·160	·707	·8106
7	·269	·760	− ·007	·6500
8	·110	·060	·696	·5001

The sums of squares of each row ought to give the same values for h^2 as did the original table in I_0, II_0, and III_0. And the inner product of any pair of rows ought to be identical also. For example, taking the last pair (it is sufficient to check adjacent rows), we have from this table :

$$\cdot 269 \times \cdot 110 + \cdot 760 \times \cdot 060 - \cdot 007 \times \cdot 696 = \cdot 0703$$

and from the other :

$$\cdot 6 \times \cdot 5 - \cdot 5 \times \cdot 3 - \cdot 2 \times \cdot 4 = \cdot 07$$

We have now succeeded in replacing our original analysis, which had many negative loadings, by one which has only positive loadings (except for the two loadings which, although negative, are nearly zero), and gives the same correlations and communalities.

4. *An orthogonal rotating matrix.*—If the reader will, in imagination, picture in his mind those original axes I_0, II_0, and III_0 as three lines at right angles to each other (orthogonal, as we say), he can further imagine them being turned bodily, using their common meeting-place as the swivelling-point and keeping them orthogonal, into their final positions I_2, II_1, and III_1. Actually we did it in two steps, but imagine it happening as one complex movement.

Arithmetically, this one movement can be imitated by "post-multiplying the original table of loadings by an orthogonal matrix," a piece of jargon we must hasten to explain. And the reader may miss this section out on first reading. A matrix, in mathematics, is an oblong or square set of numbers, to be used as an operator on other quantities. In our case it is to be used to rotate the original loadings to new positions. And since we want the axes to remain orthogonal, we have to use an orthogonal matrix, i.e. one in which the sum of the squares of any column or row is unity, and the inner product of any pair of rows or of columns is zero. Actually the orthogonal matrix which performs the rotation of the above section 3 is :

$$\begin{bmatrix} \cdot5512 & \cdot6000 & \cdot5800 \\ -\cdot4134 & \cdot8000 & -\cdot4350 \\ -\cdot7250 & \cdot0000 & \cdot6890 \end{bmatrix}$$

(The reader can check the sum of squares of any column or row, and any inner product of a pair.) Before explaining how these numbers are arrived at, let us first perform the post-multiplication of the table of original loadings (itself an oblong matrix) by this rotating matrix—

$$\begin{bmatrix} \cdot4 & \cdot4 & \cdot1 \\ \cdot7 & \cdot3 & -\cdot4 \\ \cdot7 & -\cdot2 & -\cdot3 \\ \cdot9 & -\cdot1 & \cdot3 \\ \cdot5 & \cdot2 & -\cdot2 \\ \cdot8 & -\cdot4 & \cdot1 \\ \cdot6 & \cdot5 & -\cdot2 \\ \cdot5 & -\cdot3 & \cdot4 \end{bmatrix} \times \begin{bmatrix} \cdot5512 & \cdot6000 & \cdot5800 \\ -\cdot4134 & \cdot8000 & -\cdot4350 \\ -\cdot7250 & \cdot0000 & \cdot6890 \end{bmatrix} = \begin{bmatrix} -\cdot017 & \cdot560 & \cdot127 \\ \cdot552 & \cdot660 & \cdot000 \\ \cdot686 & \cdot260 & \cdot286 \\ \cdot320 & \cdot460 & \cdot772 \\ \cdot338 & \cdot460 & \cdot065 \\ \cdot534 & \cdot160 & \cdot707 \\ \cdot269 & \cdot760 & -\cdot007 \\ \cdot110 & \cdot060 & \cdot696 \end{bmatrix}$$

We have to say *post*-multiplication because in matrix algebra the product AB is not the same as the product BA.

Matrix multiplication is performed by finding the inner product of each *row* of the first matrix with each *column* of the second matrix. Thus—

$$\cdot 4 \times \cdot 5512 - \cdot 4 \times \cdot 4134 - \cdot 1 \times \cdot 7250 = - \cdot 0174 \text{ or } - \cdot 017$$

the first item in the product matrix above. Similarly, the quantity ·707, which appears in the sixth row and third column of the product matrix, is the inner product of the sixth row of the first matrix and the third column of the second—

$$\cdot 8 \times \cdot 5800 + \cdot 4 \times \cdot 4350 + \cdot 1 \times \cdot 6890 = \cdot 7069 \text{ or } \cdot 707$$

The reader can similarly check the other entries in the product matrix.

When we performed the first of our previous two-by-two rotations we were in effect post-multiplying the loadings by the rotating matrix—

$$\begin{bmatrix} \cdot 8 & \cdot 6 & 0 \\ -\cdot 6 & \cdot 8 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

which will leave the column III_0 unchanged because of the nature of the third column of this rotating matrix. The inner product of 0, 0, and 1 with any row of the centroid loadings will give a column of loadings identical with III_0.

When we performed the second two-by-two rotation, of I_1 and III_0, we were in effect multiplying by the matrix—

$$\begin{bmatrix} \cdot 689 & 0 & \cdot 725 \\ \cdot 000 & 1 & \cdot 000 \\ -\cdot 725 & 0 & \cdot 689 \end{bmatrix}$$

which clearly does not alter the middle axis. And the rotating matrix which would have done these two operations simultaneously is—

$$\begin{bmatrix} \cdot 8 & \cdot 6 & 0 \\ -\cdot 6 & \cdot 8 & 0 \\ 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} \cdot 689 & 0 & \cdot 725 \\ \cdot 000 & 1 & \cdot 000 \\ -\cdot 725 & 0 & \cdot 689 \end{bmatrix} = \begin{bmatrix} \cdot 5512 & \cdot 6000 & \cdot 5800 \\ -\cdot 4134 & \cdot 8000 & -\cdot 4350 \\ -\cdot 7250 & \cdot 0000 & \cdot 6890 \end{bmatrix}$$

5. *Reyburn and Taylor's method.*—These South African psychologists have proposed to let psychological insight alone guide the rotations to which axes are subjected. They do not necessarily insist on a g (see their 1941*a*, pages 253, 254, 258, etc.). Their plan is to choose a group of tests which their psychological knowledge, and a study of all that is previously known, leads them to consider to be clustered round a factor. They therefore cause one of their axes to pass through the centroid of this cluster, keeping all axes orthogonal. This factor axis they do not subsequently move. They then formulate a hypothesis about a second factor and select a second group of tests, through whose centroid (retaining orthogonality) they pass their second factor axis. And so on. There is some affinity between this and Alexander's method of rotation (see page 79).

The arithmetical details of their method are as follows. They first obtain a table of centroid loadings in the usual way. Then, having chosen a group of tests which they think form, psychologically, a cluster, they add together the rows of the centroid table which refer to those tests, thus obtaining numbers proportional to the loadings of their centroid. These, after being normalized, form the first column of their rotating matrix. For example, consider this (imaginary and invented) table of loadings :

Loadings

	I	*II*	*III*	h^2
1	·4	·3	·1	·26
2	·5	—·3	—·6	·70
3	·6	—·3	—·3	·54
4	·5	·2	·1	·30
5	·4	·4	—·2	·36
6	·5	—·4	·2	·45
7	·5	·2	—·1	·30
8	·7	·4	·1	·66
9	·7	—·2	·3	·62
10	·6	—·4	·4	·68

Reyburn and Taylor now decide, let us suppose, that Tests 9 and 10 are, in their psychological view, very strongly impregnated with a verbal factor, and determine to rotate their original factors until one of them passes through the centroid of these two tests. They extract their rows, add them together, and normalize the three totals thus :

(9)	·7	—·2	·3	
(10)	·6	—·4	·4	
	1·3	—·6	·7	Sum of squares $2{\cdot}54 = 1{\cdot}594^2$
	·816	—·376	·439	obtained by dividing by 1·594

If the columns of the original table are multiplied by these three numbers and the rows added, the result is the first column of the rotated factor loadings in the table below. To get the other two columns we must complete the rotating matrix in such a manner that the axes remain orthogonal. How this is done will be explained separately later. Meanwhile, consider the matrix—

$$\begin{bmatrix} \cdot 816 & \cdot 399 & \cdot 417 \\ -\cdot 376 & -\cdot 183 & \cdot 909 \\ \cdot 439 & -\cdot 898 & . \end{bmatrix}$$

Its first column is composed of the above numbers. It is orthogonal, for the sum of the squares of any row or column is unity, and the inner product of any two is zero. When the original table of loadings is post-multiplied by this we get the rotated table :

	Rotated Loadings			h^2
1	·258	·015	·440	·260
2	·257	·793	—·064	·699
3	·471	·564	—·022	·540
4	·377	·073	·390	·300
5	·088	·266	·530	·359
6	·646	·093	—·155	·450
7	·289	·253	·390	·300
8	·465	·116	·656	·660
9	·778	·047	·110	·620
10	·816	—·047	—·113	·681

At this point the usual two checks must be made, of h^2 and of the inner products of consecutive rows.

The first factor now goes through the centroid of Tests 9 and 10, and we scan the loadings it has in the other tests to see if these are consistent with their psychological nature. For instance, Test 5 has practically no loading on this verbal factor—is this consistent with our psychological opinion of this test ?

If this scrutiny is satisfactory, the psychologist using this method then proceeds to consider where he will place his second factor ; for the second and third columns of the above loadings have still no necessary psychological meaning as they stand. Exactly the same procedure is carried out with them, the first column being left unaltered. Suppose the psychologist decided on Tests 5, 7, 8 as being a cluster round (say) a numerical factor. He adds their rows—

(5)	·266	·530	
(7)	·253	·390	
(8)	·116	·656	
	·635	1·576	
	·374	·928	when normalized

and uses their normalized totals as the first column of a matrix to rotate these last two columns. The matrix must be orthogonal, and it is in fact—

$$\begin{bmatrix} \cdot374 & \cdot928 \\ \cdot928 & -\cdot374 \end{bmatrix}$$

When the second and third columns are rotated by post-multiplication by this, the final result is given opposite. (The same checks must now be repeated.) The psychologist now scans column two to see if the loadings of his numerical factor agree reasonably with his idea of each test, and is rather sorry to see two negative loadings, but consoles himself by thinking that they are small. He must finally try to name his third factor, present to an

Final Rotated Loadings

1	·258	·414	−·151
2	·257	·237	·760
3	·471	·191	·532
4	·377	·389	−·078
5	·088	·591	·049
6	·646	−·109	·144
7	·289	·457	·089
8	·465	·652	−·138
9	·778	·120	·002
10	·816	−·122	−·001

appreciable extent only in tests 2 and 3. If he thinks he recognizes it, he is content.

6. *Special orthogonal matrices.*—To carry out the above process the reader needs to have at his disposal orthogonal matrices of various sizes, such that he can give the first column any desired values. The following will serve his purpose. Except for the first one, they are not unique, and alternatives can be made.

Order 2 $$\begin{bmatrix} u & v \\ v & -u \end{bmatrix}, \quad u^2 + v^2 = 1$$

Order 3 $$\begin{bmatrix} mq & mp & l \\ -lq & -lp & m \\ p & -q & \cdot \end{bmatrix}, \quad \left. \begin{matrix} l^2 + m^2 = 1 \\ p^2 + q^2 = 1 \end{matrix} \right\}$$

It was from this formula that the matrix used in the last section, with first column of ·816, −·376, ·439, was made. For if we set

$$\begin{aligned} p &= \cdot 439 \\ \text{we have}\quad q &= \cdot 898 \\ \text{and from } mq &= \cdot 816 \\ \text{we have}\quad m &= \cdot 909 \\ \text{and thence}\quad l &= \cdot 417 \end{aligned}$$

Order 4

$$\begin{bmatrix} a & b & -c & -d \\ b & -a & d & -c \\ c & d & a & b \\ d & -c & -b & a \end{bmatrix}$$

This one was used by Reyburn and Taylor in their 1939 article (page 159).

Similar matrices of higher order can be made by a recipe given by them, viz. multiplying together two or more of the above, suitably extended by ones and zeros. For example, a matrix orthogonal and with arbitrary first column, of order 5, can be made by multiplying together :

$$\begin{bmatrix} \cdot & \cdot & \mu\varphi & -\lambda\varphi & \pi \\ \cdot & \cdot & \mu\pi & -\lambda\pi & -\varphi \\ \cdot & \cdot & \lambda & \mu & \cdot \\ \cdot & 1 & \cdot & \cdot & \cdot \\ 1 & \cdot & \cdot & \cdot & \cdot \end{bmatrix} \times \begin{bmatrix} mq & -lq & p & \cdot & \cdot \\ mp & -lp & -q & \cdot & \cdot \\ l & m & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & 1 & \cdot \\ \cdot & \cdot & \cdot & \cdot & 1 \end{bmatrix}$$

where $l^2 + m^2 = p^2 + q^2 = \lambda^2 + \mu^2 = \pi^2 + \varphi^2 = 1$.

7. *Principles deciding where to stop rotation.*—We have mentioned two principles, (*a*) the desire to rotate to positions where there will be few, if any, negative loadings —but usually this is insufficient to define a final position uniquely, and (*b*) Reyburn and Taylor's plan of following their psychological intuition in placing the axes. They too accept the need for mainly positive loadings, and they keep their axes at right angles. We turn now, in our next chapter, to a principle (Simple Structure) which is accepted widely in America, though hardly at all in Great Britain.

CHAPTER XI

ORTHOGONAL SIMPLE STRUCTURE

1. *Agreement of mathematics and psychology.*—It is clear that the whole process of multifactor analysis is one by which a definition of the primary factors is arrived at by satisfying simultaneously certain mathematical principles and certain psychological intuitions. When these two sides of the process click into agreement, the worker has a sense of having made a definite step forward. The two support one another. Obviously the goal to be hoped for along this line of advance will be the discovery of some mathematical process which always leads to a unique set of factors mainly acceptable to the psychologist. If such could be discovered and found to produce a few factors over and above those recognized as already known by other means, the new factors would stand a good chance of acceptance on the strength of their mathematical descent only. And no doubt the psychologist would be prepared to make a few concessions and changes in his previous ideas to fit in with any mathematical scheme which already gave much satisfaction and was objective and unique in its results.

It is here that Thurstone's notion of " simple structure " is offered as a solution (*Vectors*, Chapters 6–8). This idea is that the axes are to be rotated until as many as possible of them are at right angles to as many as possible of the original test vectors ; and that the battery is not suitable for defining factors unless such a rotation is uniquely possible, a rotation which will leave every axis at right angles to at least as many tests as there are factors, and every test at right angles to at least one axis.

When the vectors of a test and a factor are at right angles, the loading of the factor in that test is zero. Thurstone's " simple structure " is therefore indicated by a large number of zeros in the matrix of loadings, so large that there will be only one position of the axes (if any) which satisfies the requirement. His search, be it repeated, is for a set of conditions which will make the solution unique. We have seen him approaching this goal by stages. Unless the battery is large, so that—

$$n \geqslant \frac{(2r + 1) + \sqrt{(8r + 1)}}{2}$$

(see Chapter V, Section 9), the communalities are not unique. Even when the battery is large enough, the axes representing factors may be rotated to positions among which there is no one specially marked out. Then comes the demand that there be this large number of zero loadings. Most batteries of tests will not allow this demand to be satisfied, but with some it can just be attained. Only these last, it is Thurstone's conviction, are suitable for defining primary factors, and it is his faith that the factors thus mathematically defined will be found to be acceptable as psychologically separable unitary traits.

2. *An example of six tests of rank* 3.—To make our remarks more definite and concrete, let us suppose that we have a battery of six tests whose matrix of correlations can be reduced to rank 3. In practice, of course, six tests are far too few, and more than three factors quite likely. The matrix of loadings given by the " centroid " system contains at first negative quantities. Thus from the correlations :

	1	2	3	4	5	6
1	.	·525	·000	·000	·448	·000
2	·525	.	·098	·306	·349	·000
3	·000	·098	.	·133	·314	·504
4	·000	·306	·133	.	·000	·000
5	·448	·349	·314	·000	.	·307
6	·000	·000	·504	·000	·307	.

with the communalities—

·674	·634	·558	·415	·490	·493

we get by the " centroid " process the matrix of loadings :

	I	*II*	*III*
1	·542	·612	·074
2	·629	·342	—·348
3	·529	—·492	·191
4	·281	—·182	—·550
5	·628	·143	·274
6	·429	—·424	·359

It is the factor axes indicated by these loadings that Thurstone wishes to rotate until there are no negative loadings and enough zero loadings to make the position uniquely defined. For this last purpose he finds, empirically, that it is necessary to require—

(*a*) At least one zero loading in each row ;

(*b*) At least as many zero loadings in each column as there are columns (here three) ; and

(*c*) At least as many *XO* or *OX* entries in each pair of columns as there are columns. By an *XO* entry is meant a loading in the one column opposite a zero in the other.

" At least one zero loading in each row." This means that no test may contain all the common factors. In making up the battery, then, the experimenter, with some idea in his mind as to what the factors are, will endeavour to ensure that they are not all present in any one test. This would, for example, exclude from a Thurstone battery (except as an extra) any very mixed group test, or a mixed test like the Binet-Simon which is itself a whole battery of varied items.

" At least as many zeros in each column as there are columns," that is, as there are common factors. This means that in a Thurstone battery no factor may be general, but must be missing in several tests.

The requirement as to the number of *XO* or *OX* entries is intended to ensure that the tests are qualitatively distinct from one another.

Now, these requirements cannot generally be met by a matrix of loadings. It will in general be impossible to rotate the axes (keeping them orthogonal) until every

axis is at right angles to r test vectors. The above a... ...ny)
example has, however, been constructed so that th... ...ed,
be done.

The correlations were in fact made from the loading...

	A	B	C
1	.	.	·821
2	.	·475	·639
3	·718	·206	.
4	.	·644	.
5	·438	.	·546
6	·702	.	.

and the centroid loadings must therefore be capable of being rotated rigidly into this form, retaining orthogonality.

3. *Two-by-two rotation to simple structure.*—The problem for the experimenter, however, is to *discover* this "simple structure," if it exists; he is not, like us, in the position of knowing that it does exist, and what it is. Thurstone's original method was to use two-by-two rotations, in each rotation endeavouring to obtain some zero loadings. Let us illustrate by our artificial example, taking first the centroid factors I and II. Using their centroid loadings as co-ordinates, we obtain Figure 23. At once we notice that the test points 3, 4, and 6 are almost collinear on a radius from the origin, and that if we rotate the axes clockwise through about 42° the new position of I, labelled I_1 in the diagram, will almost pass through these test points, while the new axis II_1 will almost pass through test point 1. On these new axes, therefore, Tests 3, 4, and 6 will have hardly any projections on axis II_1; that is, will have hardly any loadings in a factor along II_1. From tables we find sin 42° = ·669, and cos 42° ≑ ·743. We have then:

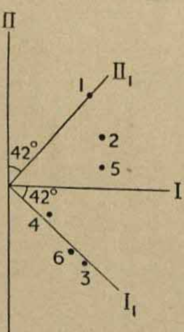


Figure 23.

	Old loadings		*New loadings*	
	I	*II*	I_1	II_1
1	·542	·612	—·007	·817
2	·629	·342	·239	·675
3	·539	—·492	·722	—·012
4	·281	—·182	·331	·053
5	·628	·143	·371	·526
6	·429	—·424	·602	—·028
multipliers	·743	—·669 for I_1 loadings,		
	·669	·743 for II_1 loadings.		

We have now obtained our desired three zero (or near zero) loadings in factor II_1. Accepting the approximations to zero as good enough for the present, we next make Figure 24 from the loadings of I_1 and III in the same way as we made the former figure. In this, Test 1 falls quite near the origin. Tests 5 and 6 are approximately on one radius, and Tests 2 and 4 on another, and these radii are at right angles to one another. If we rotate the axes I_1 and III rigidly through a clockwise turn of about 49° they will pass almost through these radial groups and nearly zero projections will result.* Using sin 49° = ·755 and cos 49° = ·656 we perform a similar calculation to the preceding, using the loadings I_1 and III as starting-point and obtaining loadings on I_2 and III_1 (the subscript indicating the number of rotations that axis has undergone). We have finally, putting our results together, the table of loadings overleaf FΛ.†

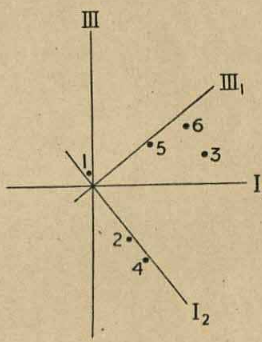


Figure 24.

* The rotation might with advantage have been carried a little further.

† The matrix symbols, using Thurstone's notation, are given for the convenience of mathematical readers. Others should ignore them. When the tests are many and the centroids few, a saving can be effected by picking tests equal in number to the factors and per-

	I_2	II_1	III_1
1	−·060	·817	·043
2	·420	·675	−·048
3	·329	−·012	·670
4	·632	·053	−·111
5	·037	·526	·460
6	·124	−·028	·690

Clearly, this is an approximation to the loadings of the factors *A*, *B*, and *C* which we who are in the secret (as a real experimenter is not) know to have been used in making the correlations : III_1 here is *A*, I_2 here is *B*, and II_1 is *C*. The small loadings are not quite zero, and the other loadings not quite the same, but a further set of rotations would refine the results and bring them nearer to the *A B C* values.

4. *New rotational method.*—When this two-by-two rotational method is used on a large battery of tests, with perhaps six or seven factors instead of three, it is not only laborious but somewhat difficult to follow. Thurstone has, however, devised a method of rotation which takes the factors three at a time, and to this we now turn, still using our small artificial example as illustration. In this example, since there are only three factors, this new method leads to a complete solution at once. With more factors the matter would be more complicated.

If the reader will think of the three centroid factors as represented by imaginary lines in the room in which he is sitting (Figure 25), he will be aided in following the explanation of this new method. Imagine the first centroid axis to be vertically in the middle of the room, and the other two centroid axes on the carpet, at right angles to the first and to each other. The test points are in various positions in the room space, if we take their three centroid loadings as co-ordinates and treat the distance from floor to ceiling as unity. Imagine each test point joined

forming two-by-two rotations on their loadings F_s. Let the resulting loadings be V_s. Then $R = F_s^{-1} V_s$ can be used as a rotating matrix on the whole table F of centroid loadings. The tests chosen to form F_s should represent different clusters.

by a line to the origin (in the middle of the carpet, where the axes cross). The lengths of these lines are the square roots of the communalities, and the loadings on the first centroid factor are their projections on to the vertical axis, the height, that is, of each test point above the floor.

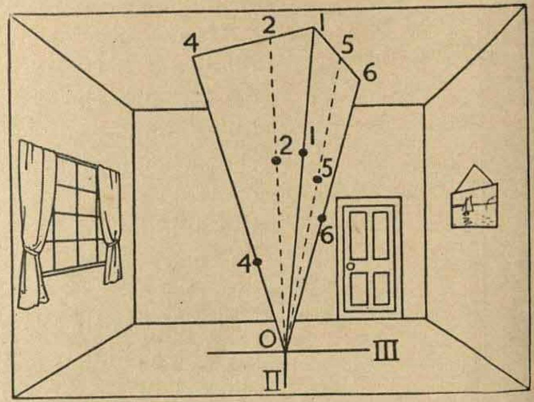


Figure 25 (not to scale).

Thurstone now imagines each of these lines or communality vectors produced until it hits the ceiling, making a pattern of dots on the ceiling. These extended vectors now all have unit projection on the first centroid axis, for we agreed to call the distance from floor to ceiling unity. Their y and z co-ordinates on the ceiling will be correspondingly larger than their loadings on the second and third centroid factors, and can be obtained by dividing each row of the centroid loadings by the first loading. In our case this gives us the following table, obtained in the manner just mentioned from the table on page 153.

Extended centroid projections

	I_e	II_e	III_e
1	1·000	1·129	·137
2	,,	·544	−·553
3	,,	−·930	·361
4	,,	−·648	−1·957
5	,,	·228	·436
6	,,	−·988	·837

The second and third columns are now the co-ordinates of those dots on the ceiling of which we spoke. A diagram of the ceiling, seen from above, is given in Figure 26 and the important point about it is that the dots form a triangle.

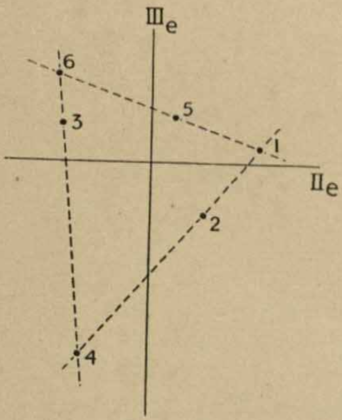


Figure 26.

If the reader will now picture this triangle as drawn on the ceiling of his room, and remember that the origin, where the centroid axes crossed, is in the middle of the carpet, he can next imagine an inverted three-cornered pyramid, with the triangle on the ceiling as its base, the origin in the middle of the carpet as its apex and the communality vectors 1, 4, and 6 as its edges. The vector 5 lies on one of the faces of this pyramid ; vector 2 lies on another ; vector 3 lies on the remaining face, all springing from the origin and going up to the ceiling.

5. *Finding the new axes.*—If now we choose for new axes (in place of the centroid axes) three lines at right angles respectively to the three plane faces of our pyramid, the test projections on these axes will clearly have the zeros we desire. The three vectors 1, 2, and 4 all lie in one face, and will have zero projections on the axis A' at right angles to that face. The vectors 1, 5, and 6 will have zero projections on the line B' at right angles to their face. The vectors 3, 4, and 6 will have zero projections on C' at right angles to their face. The reader should visualize these new axes in his room. It remains to be shown how the other, non-zero, projections are to be calculated, and to inquire whether these new axes are orthogonal, and whether they can be identified with the original A, B, and C. The first step is to obtain the equations of the three sides of the triangle in the diagram. Where there are many tests and the dots are not perfectly

collinear, one plan is to draw a line through them by eye, and measure the distances a and b it cuts off on the axes, then using the equation—

$$\frac{y}{a} + \frac{z}{b} = 1.$$

Or we can write down the equations of the lines joining points at the corners, either actual test points, or the places where our lines intersect, using the equation—

$$(lv - mu) + (m - v)\,y + (u - l)\,z = 0$$

when l, m are the co-ordinates of one corner, and u, v of another. We obtain in our case—

$$\begin{aligned} -\,2{\cdot}121 + 2{\cdot}094y - 1{\cdot}777z &= 0 \text{ for line 1, 2, 4} \\ -\,1{\cdot}080 + \ {\cdot}700y + 2{\cdot}117z &= 0 \quad ,, \quad ,, \quad 1, 5, 6 \\ 2{\cdot}476 + 2{\cdot}794y + \ {\cdot}340z &= 0 \quad ,, \quad ,, \quad 4, 3, 6 \end{aligned}$$

where y means the extended II, and z the extended III. Before we go further we have to divide each equation through by the root of the sum of the squares of its coefficients, so that the new coefficients sum to unity when squared—this is called normalizing and is necessary in order to keep the communalities right and for other reasons. The equations then are :

$$\begin{aligned} -\,{\cdot}611 + {\cdot}603y - {\cdot}512z &= 0 \quad (1) \\ -\,{\cdot}436 + {\cdot}283y + {\cdot}854z &= 0 \quad (2) \\ {\cdot}660 + {\cdot}745y + {\cdot}091z &= 0 \quad (3) \end{aligned}$$

and it is clear, from the way in which they have been reached, that these equations will be satisfied by the extended co-ordinates of certain of the rows in the table on page 153. Consider the first equation and write its co-

	—·611 x	·603 y	—·512 z	*Weighted sum*
1	·542	·612	·074	·000
2	·629	·342	—·348	·000
3	·529	—·492	·191	—·718
4	·281	—·182	—·550	·000
5	·628	·143	·274	—·438
6	·429	—·424	·359	—·701

efficients above the columns of that table, placing $-\cdot611$ over the first column, as shown at the foot of page 159. If we multiply each column by the multiplier above it and add the rows we get the quantities shown on the right for comparison with page 154. The zeros are in the right places for factor *A*. The other loadings are, however, negative, but that can be easily put right by changing all the signs of the multipliers, which we are at liberty to do. Similarly, using eqns. (2) and (3) we get the loadings of factors *B* and *C* exactly, except for an occasional difference due to rounding off at the third decimal place. We have, indeed, found the matrix product FΛ,

$$\begin{bmatrix} \cdot542 & \cdot612 & \cdot074 \\ \cdot629 & \cdot342 & -\cdot348 \\ \cdot529 & -\cdot492 & \cdot191 \\ \cdot281 & -\cdot182 & -\cdot550 \\ \cdot628 & \cdot143 & \cdot274 \\ \cdot429 & -\cdot424 & \cdot359 \end{bmatrix} \begin{bmatrix} \cdot611 & \cdot436 & \cdot660 \\ -\cdot603 & -\cdot283 & \cdot745 \\ \cdot512 & -\cdot854 & \cdot091 \end{bmatrix} = \begin{bmatrix} . & . & \cdot821 \\ . & \cdot475 & \cdot639 \\ \cdot718 & \cdot206 & . \\ . & \cdot644 & . \\ \cdot438 & . & \cdot546 \\ \cdot702 & . & . \end{bmatrix}$$

except, as has been already said, for occasional discrepancies in the third decimal place. The procedure we have described has enabled us to discover this last matrix, with which, in fact, we began. And by analogy (is the deduction sound ?) an experimenter with experimental data who follows this procedure and reaches simple structure concludes that that is how his correlations were made. Certainly that is how they *may* have been made.

The matrix Λ beginning with $\cdot611$ is the rotating matrix which turns the axes I, II, III into the new positions *A*, *B*, *C*. Its columns are the direction-cosines of *A*, *B*, and *C* with reference to the orthogonal system I, II, III. Are *A*, *B*, and *C* orthogonal ? The cosines of the angles between them can by a well-known rule be found by premultiplying the rotating matrix by its transpose. When we do so we find $\Lambda'\Lambda = \mathrm{I}$, viz. :

$$\begin{bmatrix} \cdot611 & -\cdot603 & \cdot512 \\ \cdot436 & -\cdot283 & -\cdot854 \\ \cdot660 & \cdot745 & \cdot091 \end{bmatrix} \begin{bmatrix} \cdot611 & \cdot436 & \cdot660 \\ -\cdot603 & -\cdot283 & \cdot745 \\ \cdot512 & -\cdot854 & \cdot091 \end{bmatrix} = \begin{bmatrix} 1 & . & . \\ . & 1 & . \\ . & . & 1 \end{bmatrix}$$

(again allowing for third decimal place discrepancies). That is to say, the angles between *A*, *B*, and *C* have zero cosines, they are right angles.

The axes *A*, *B*, and *C* were drawn at right angles to the three planes which form the pyramid mentioned above, and therefore these three planes are also at right angles to one another. (Our rough sketch in Figure 25 made the pyramid too acute.) It follows that *A*, *B*, and C are actually the edges of the pyramid. In our example (though this need not be the case) they happen to pass each through a test point in the room, *A* through Test 6, *B* through Test 4, and *C* through Test 1. These tests are not identical with the factors, for each test contains a specific element, not in the common-factor space, but at right angles to it. What we have called a test point is the end of the unit test vector projected on to the common-factor space. The complete test vectors are out in a space of more dimensions, of which the three-dimensional common-factor space is a subspace.

6. *Landahl preliminary rotations.*—When there are more than three centroid factors, the calculations are not so simple. If the common-factor space is, for example, four-dimensional, then the table of extended vectors, in addition to its first column of unities, will have three other columns. The two-dimensional ceiling of our room, in our former analogy, has here become three-dimensional, a *hyper*-plane at right angles to the first centroid axis. On paper its dimensions can only be graphed two at a time, and no complete triangle will be visible among the dots. But sets of dots will be seen to be collinear, lines can be drawn through them, and a procedure similar to that outlined above followed. This will become clearer when we work a four-dimensional example. First, however, it is desirable to explain, on our simple three-dimensional example, a device which facilitates the work on higher dimensional problems, called the Landahl rotation. It is unnecessary in the three-dimensional case, and we are using it only to explain it for use with more than three dimensions.

A Landahl rotation turns the centroid axes solidly to a position where each of them is equally inclined to the original first centroid axis. In our imagined room the first centroid axis ran vertically from the middle of the floor to the middle of the ceiling, while the other two were drawn on the floor itself. Imagine all three (retaining their orthogonality) to be moved, on the origin as pivot, until they are equally inclined to the vertical so that they enclose the inverted pyramid of Figure 25. That is a Landahl rotation. The lines through the test points have not moved. They remain where they were, and still hit the ceiling in the same pattern of dots. The projections of the extended vectors on to the original first centroid axis all still remain unity. But for the next step in this method we need their projections on to the Landahl axes. We obtain these by post-multiplying the matrix of centroid extended loadings by a Landahl matrix, an orthogonal matrix with each element in its first row equal to $\frac{1}{\sqrt{c}}$, where c is the order of the matrix; that is, its number of rows or columns (Landahl, 1938). We need a Landahl matrix of order 3, for example:

$$\begin{bmatrix} \cdot 577 & \cdot 577 & \cdot 577 \\ \cdot 816 & -\cdot 408 & -\cdot 408 \\ \cdot 000 & \cdot 707 & -\cdot 707 \end{bmatrix}$$

The element ·577 is the cosine of the angle which each axis makes, after rotation, with the original position of the first centroid axis.

When the table of extended vector projections on page 157 is post-multiplied by the above matrix, the table on page 163 results, giving the projections of the extended vectors on to the Landahl axes L, M, N.

From this table three diagrams LM, LN, and MN can be made, and the reader is advised to draw them. Each of them shows a triangular distribution of dots and in this simple three-dimensional example only one of them is needed. But in a multi-dimensional problem several are

*Projections on Landahl axes**

	L	M	N
1	1·498	·213	·020
2	1·021	—·066	·746
3	—·182	1·212	·701
4	·048	—·542	2·225
5	·760	·792	·176
6	—·229	1·572	·388

needed, and as a rule only one line is used on each diagram employed. Here, from the LN diagram we find the equations of the three sides of the triangle to be :

$$\begin{aligned} -2{\cdot}205l - 1{\cdot}450n + 3{\cdot}332 &= 0 \\ {\cdot}368l + 1{\cdot}727n - {\cdot}586 &= 0 \\ 1{\cdot}837l - {\cdot}277n + {\cdot}528 &= 0 \end{aligned}$$

We want to make these homogeneous in l, m, and n, and so we add, after each of the numerical terms, the factor $\cdot 577\ (l + m + n)$, which equals unity. The equations then are :

$$\begin{aligned} -{\cdot}282l + 1{\cdot}923m + {\cdot}473n &= 0 \\ {\cdot}030l - {\cdot}338m + 1{\cdot}389n &= 0 \\ 2{\cdot}142l + {\cdot}305m + {\cdot}028n &= 0 \end{aligned}$$

* After a Landahl adjustment the axes are not infrequently already near simple structure, as here. It is sometimes worth while to rotate them slowly round the original first centroid, like spinning an umbrella, to improve the approximation to zero entries. This can be done by an orthogonal matrix whose columns sum to unity, as e.g.

$$\begin{bmatrix} {\cdot}9900 & -{\cdot}0946 & {\cdot}1046 \\ {\cdot}1046 & {\cdot}9900 & -{\cdot}0946 \\ -{\cdot}0946 & {\cdot}1046 & {\cdot}9900 \end{bmatrix}$$

or its transpose : and the rotation will be the slower, the nearer the diagonal elements are to unity.

After normalizing, these become :

$$-\cdot141l + \cdot961m + \cdot236n = 0$$
$$\cdot021l - \cdot236m + \cdot971n = 0$$
$$\cdot990l + \cdot141m + \cdot013n = 0$$

Writing the coefficients as columns in a matrix, and premultiplying by Landahl's matrix (since at an earlier stage we post-multiplied by it) we obtain :

$$\begin{bmatrix} \cdot609 & \cdot436 & \cdot660 \\ -\cdot603 & -\cdot283 & \cdot745 \\ \cdot513 & -\cdot853 & \cdot090 \end{bmatrix}$$

the same matrix Λ as we arrived at (page 160) without the use of Landahl's rotation. The advantage of using a Landahl rotation appears only in problems with more than three common factors. The reader can readily make a Landahl matrix of any required order, say 5. Fill the first row with the root reciprocal of 5, ·447. Complete the first column by putting in the second place ·894 (because $\cdot447^2 + \cdot894^2 = 1$), and below that zeros. The second row must then be completed with equal elements, all negative, such that the row sums to zero. Then the second column is completed in a similar way, and the third row, and so on. The reader should finish it. There are alternative forms possible, one of which is used below.

An unfinished Landahl matrix :

$$\begin{bmatrix} \cdot447 & \cdot447 & \cdot447 & \cdot447 & \cdot447 \\ \cdot894 & -\cdot224 & -\cdot224 & -\cdot224 & -\cdot224 \\ \cdot000 & \cdot866 & -\cdot289 & -\cdot289 & -\cdot289 \\ \cdot000 & \cdot000 & & & \\ \cdot000 & \cdot000 & & & \end{bmatrix}$$

7. *A four-dimensional example.*—The following example of a problem with four common factors is only partly worked out, so that the reader can finish it as an exercise. It also is an artificial example, and orthogonal simple structure can be arrived at. The centroid analysis gave four centroid factors with the loadings shown in this table :

Centroid loadings F

	I	*II*	*III*	*IV*
1	·727	·517	·094	·126
2	·575	·105	·553	·049
3	·810	·289	·246	—·246
4	·588	·417	—·367	—·382
5	·524	—·583	—·450	·183
6	·549	—·435	·398	—·013
7	·624	—·318	—·187	—·254
8	·594	—·551	·239	·084
9	·626	·252	—·169	·562
10	·645	·307	—·357	—·109

After these have been " extended " (i.e. divided in each row by the first loading) they were post-multiplied by a Landahl matrix, one of the alternative forms, viz. :

$$\begin{bmatrix} \cdot5 & \cdot5 & \cdot5 & \cdot5 \\ \cdot5 & \cdot5 & -\cdot5 & -\cdot5 \\ \cdot5 & -\cdot5 & \cdot5 & -\cdot5 \\ \cdot5 & -\cdot5 & -\cdot5 & \cdot5 \end{bmatrix}$$

and the resulting projections on the Landahl axes were thus found to be :

	L	*M*	*N*	*P*
1	1·007	·704	·122	·166
2	1·115	·068	·848	—·030
3	·679	·678	·625	·018
4	·218	1·492	·158	·132
5	—·311	·199	·453	1·660
6	·455	—·247	1·270	·522
7	—·107	·598	·808	·701
8	·308	—·235	1·094	·833
9	1·015	·387	—·285	·883
10	·376	1·099	·070	·454

Six diagrams can be made, and it is advisable to draw them all, though not all are necessary. The *LN* diagram is shown in Figure 27. We scan it for collinear points (not necessarily radial) *which have all or nearly all the other* *points on one side of their line*, and note the line 5, 4, 10, 9. Its equation is readily found to be approximately :

$$\cdot738l + 1\cdot327n - \cdot371 = 0.$$

We make this homogeneous by substituting for unity, after the numerical term ·371, the quantity $\cdot5\ (l + m + n + p)$, for ·5 is the cosine of the angle each of the Landahl axes makes with the original first centroid axis. This gives us the equation (not yet normalized) :

$$\cdot553l - \cdot185m + 1\cdot141n - \cdot185p = 0.$$

Three more equations are needed, and one of them can indeed be obtained from the same diagram, on which points 5, 7, 8, 6 are very nearly collinear. The reader is advised to draw the remaining diagrams and complete the calculations following the steps of our previous example. The above equation refers to a line which makes a fairly big angle with N. It is desirable to look for the remaining three lines making large angles (approaching right angles) with L, M, and P.

It will be remembered that in our earlier example the sign of one equation had to be changed at the end of the calculation because large negative values were appearing in the final matrix of loadings. This can be obviated by attending to the following rule. If the other test-points are on the same side of the line as the origin the numerical term must be positive in the equation ; if they are on the side remote from the origin the numerical term must be negative. In the adjacent diagram, the origin and the other points are on opposite sides of the line through 5, 4, 10, 9 and therefore the numerical term must be negative, as it is (—·371). Had it been positive all the signs of the equation would have required to be changed.

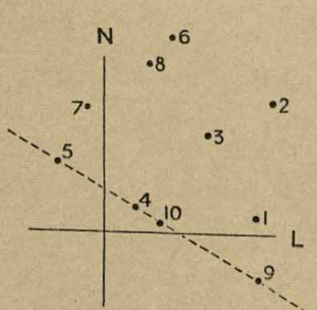


Figure 27.

8. *Ledermann's method of reaching simple structure.*— Ledermann has pointed out that when simple structure

can be attained (whether orthogonal or oblique) then as many r-rowed principal minors of the reduced correlation matrix must vanish as there are common factors ; and that it follows that the same number of vanishing determinants must be discoverable in the table of centroid loadings. Thus, for example, in the table of centroid loadings on page 153 the three determinants composed respectively of rows 1, 2, and 4 ; of rows 1, 5, and 6 ; and of rows 3, 4, and 6 all vanish, and these rows are where the zeros come in the three columns of the simple structure. This gives an alternative method of reaching simple structure. Test every possible r-rowed determinant in the centroid table of r factors. If r of them are discovered to vanish, then simple structure may be and probably is possible. Each of these vanishing determinants will provide a column of the rotating matrix Λ, for which purpose we delete any one of its rows and calculate all the $r-1$ rowed minors from what is left. The column has then to be normalized. This process works equally well for oblique simple structure (see next Chapter). Its drawback, when the number of factors is large, is the necessity of calculating so many determinants to discover those that vanish.

9. *Limits to the extent of factors.**—Orthogonal simple structure requires that no factor shall extend through many tests, and it is possible to decide beforehand, from the correlations, whether factors running through not more than s tests each are adequate to give the measured correlations, leaving $n - s$ zeros. They will not as a rule be able to do so if the average correlation exceeds $(s - 1)/(n - 1)$: more exactly, not if the largest latent root of the matrix is larger than s. If these rules are to be applied when communalities are used, as is the case when testing whether orthogonal simple structure is possible, the matrix should first be " corrected for communality," i.e. each r must be divided by the square root of the product of the two communalities concerned. Approximations to the largest latent root of a matrix of correlations, when the entries are all positive, are—

* A brief summary of a chapter with this title in previous editions.

$$\frac{\text{sum of the whole matrix}}{n}$$

or more accurately—

$$\frac{\text{sum of the squares of the column totals}}{\text{sum of the whole matrix}}$$

An exact test for the possibility of orthogonal simple structure has been given (Ledermann, 1936) and is described in the Appendix, page 367, but it requires a prohibitive amount of calculation.

Even, however, when orthogonal simple structure cannot be attained with orthogonal factors, it may be possible to reach it with oblique factors.

10. *Leading to oblique factors.*—In this chapter we have kept our factors orthogonal; that is, independent, uncorrelated with one another. It is natural to desire them to be *different* qualities, and convenient statistically. In describing a man, or an occupation, it would seem to be both confusing and uneconomical to use factors which, as it were, overlapped. Yet in situations where more familiar entities are dealt with, we do not hesitate to use correlated measures in describing a man. For instance, we give a man's height and weight, although these are correlated qualities.

Often, moreover, a battery of tests which will not permit simple structure to be reached if orthogonal factors are insisted on will nevertheless do so if the factors are allowed to sag away a little from strict orthogonality. Even as early as in *Vectors of Mind*, Thurstone expressly permitted this. It can clearly be defended on the ground that even if the factors were uncorrelated in the whole population, they might well be correlated to some extent in the sample of people actually tested. I was at one time under the impression that this comparatively slight departure from orthogonality was all that was contemplated by Thurstone. But he and his fellow-workers now have the courage of their convictions, and permit factors to depart from orthogonality as much as is necessary to attain simple structure, even if they are then found to be quite highly correlated. A chapter on these oblique factors* is therefore necessary, and out of them arise Thurstone's "second order factors."

11. *Parallel proportional profiles.*—A method which, like Thurstone's simple structure, is meant to enable us to arrive at factors which are real entities, or to check whether our hypotheses about the factor composition of tests are correct, has been put forward by R. B. Cattell (1944*b*, 1946), and has interesting possibilities which its author will no doubt develop. The essence of his idea is that "if a factor is one which corresponds to a true functional unity, it will be increased or decreased 'as a whole'," and therefore if the same tests are given under two different sets of circumstance, which favour a certain factor more in one case and less in the other, the loadings of the tests in that factor should all change in the same proportion. Experimental trials of this principle may be expected soon from its author. Among "different circumstances" he mentions different samples of subjects, differing, say, in age or sex, and different methods of scoring, or different associated tests in the battery. But he prefers another kind of change of circumstance; namely a change "from measures of static, inter-individual differences to measures from other sources of differences in the same variables." He instances, among his examples, intercorrelating changes in scores of individuals with time, or intercorrelating differences of scores in twins. We may thus have two, or several, centroid analyses, and the mathematical problem is to find rotations which will leave the profile of loadings of a certain factor similar in all the factor matrices. It may even be that the profiles of several factors could be made similar. These factors would then satisfy Cattell's requirement as corresponding to "true functional unities." The necessary modes of calculation to perform these rotations have not yet been more than adumbrated, however.

* It must be clearly understood that this obliquity or correlation of factors is quite a different matter from the correlation of *estimates*, even of orthogonal factors, due to the excess of factors over tests described on pages 237 to 242.

CHAPTER XII

OBLIQUE FACTORS

1. *Pattern and structure.*—So long as the factors are orthogonal, the loadings in the matrix of loadings are also the correlations between the factor and the tests, but this ceases to be the case when the factors are correlated. The word "loading" continues to be used for the coefficients such as l, m, and n in equations like—

$$z = l\alpha + m\beta + n\gamma$$

and the matrix or table of these is called a *pattern*, while the matrix of correlations between tests and factors is called a *structure*. The entries in a *structure* are projections from a point on to certain axes. The entries in a *pattern* are the oblique co-ordinates of that point along those axes. The two are only identical if the axes are orthogonal.

Moreover, as soon as the factors become oblique, it becomes necessary to distinguish between "reference vectors" and "primary factors." The reference vectors are the positions to which the centroid axes have been rotated so that the test-projections on to them include a number of zeros. Each reference vector is at right angles to a hyperplane containing a number of communality vectors. A hyperplane is a space of one dimension less than the common-factor space. In our first example in Chapter XI the hyperplanes were ordinary planes, the faces of the three-cornered pyramid there referred to (see page 157) and each reference vector was at right angles to one of those faces.

The primary factor corresponding to a given reference vector is the line of intersection of all the *other* hyperplanes, excluding, that is, the hyperplane at right angles to the reference vector. In our three-dimensional common-factor space the primary factor was the edge of the pyramid where those two faces met, excluding that face to which the reference vector was orthogonal.

Now, when the reference vectors turn out to be at right angles to each other, as they did in that example, each reference vector is identical with its own primary factor. But not when the reference vectors turn out to be oblique. In Chapter XI we did not distinguish them, and called their common line the " factor." But in this chapter the distinction must be kept clearly in mind. It is the primary factors Thurstone wants. The reference vectors are only a means to an end.

Thurstone's second method of rotation described in Chapter XI, the method in which the communality vectors are " extended," and lines drawn on the diagrams which are not necessarily radial lines, will not keep the axes orthogonal, but seeks for the axes on which a number of projections are zero, regardless of whether the resulting directions are orthogonal or oblique. *In general they will be oblique*, and the examples worked in Chapter XI only gave *orthogonal* simple structure because they had been devised so as to do so. The test of orthogonality is that the matrix of rotation, premultiplied by its transpose, gives the unit matrix (see page 160). Or in other words, that the inner products of the columns of the rotating matrix are all zero. They are the cosines of the angles between the reference vectors, and the cosine of 90° is zero.

2. *Three oblique factors.*—To illustrate Thurstone's method when the resulting factors are oblique we shall next work an example devised to give three oblique common factors. Consider this matrix of correlations :

	1	2	3	4	5	6	7
1		·728	·167	·372	·153	·105	·126
2	·728		·696	·583	·651	·347	·638
3	·167	·696		·857	·775	·709	·740
4	·372	·583	·857		·543	·797	·473
5	·153	·651	·775	·543		·504	·828
6	·105	·347	·709	·797	·504		·433
7	·126	·638	·740	·473	·828	·433	

which, with guessed communalities, gives these centroid loadings :

	F		
	I	*II*	*III*
1	·449	—·682	·165
2	·825	—·478	—·129
3	·906	·336	·020
4	·846	·133	·457
5	·808	·208	—·412
6	·697	·336	·335
7	·767	·173	—·468

When these projections on the centroid axes are " extended," that is, when each row is divided by the first loading in that row, we obtain this table :

	I_e	II_e	III_e
1	1·000	—1·519	·367
2	,,	—·579	—·156
3	,,	·371	·022
4	,,	·157	·540
5	,,	·257	—·510
6	,,	·482	·481
7	,,	·226	—·610

The columns II_e and III_e in this table represent the co-ordinates of the " dots on the ceiling " in our analogy of Chapter XI, p. 157 . When we make a diagram of them we obtain Figure 28. We see that a triangular formation is present, and we draw the dotted lines shown.

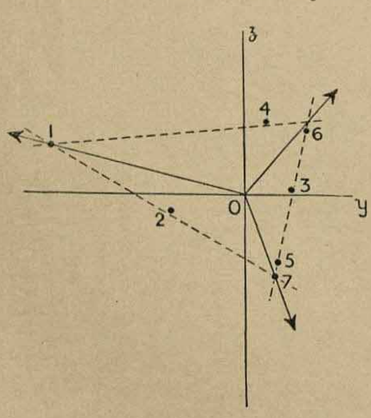


Figure 28.

It is not essential, it may be remarked in passing, that there be no points elsewhere than on the lines, provided they are additional to those required to fix the simple structure. Had it not been for the desirability of keeping the example small we would have increased the number of tests, and not only arranged for further points to fall on these lines, but also included some whose dots fell inside the triangle, representing tests which involve all three factors.

We find the equations of these lines to be approximately

$$\begin{aligned}
\cdot 475 + \ \cdot 50y \ + \ \cdot 95z &= 0 \text{ (line 1, 2, 7)} \\
1\cdot 113 + \ \cdot 183y - 2\cdot 119z &= 0 \text{ (line 1, 4, 6)} \\
\cdot 403 - 1\cdot 091y + \ \cdot 256z &= 0 \text{ (line 7, 5, 3, 6)}
\end{aligned}$$

The coefficients of each equation have to be "normalized," that is, reduced proportionately so that the sum of their squares is unity (for they are to be direction cosines). These normalized coefficients are then written as columns in a matrix as follows :

$$\begin{bmatrix} \cdot 405 & \cdot 464 & \cdot 338 \\ \cdot 426 & \cdot 076 & -\cdot 916 \\ \cdot 809 & -\cdot 883 & \cdot 215 \end{bmatrix} = \Lambda$$

The table of centroid loadings on page 172 must now be post-multiplied by this rotating matrix to obtain the projections of the tests on the three *reference vectors* which are at right angles to the planes defined by the dotted lines in our diagram. We obtain this table :

$$V = F\Lambda$$

(Simple) Structure on the Reference Vectors

	L'	B'	D'
1	·025	·011	·812
2	·026	·460	·689
3	·526	·428	·003
4	·769	—·001	·262
5	·083	·755	—·006
6	·696	·053	·000
7	·006	·782	·000

We have labelled the columns L', B', and D' for a reason which will become apparent later, when we explain how the correlations were, in fact, made. This table is a simple structure, formed by the projections on the reference vectors. It has a zero (or near-zero) in each row, and three or more in each column, in the positions to be anticipated from Figure 28; for example, tests 3, 5, 6, and 7, which are collinear in the figure, have zeros in column D'.

Now let us test the angles between the reference vectors. To do this we premultiply the rotating matrix by its transpose

$$\Lambda'\Lambda = \mathrm{C}$$

$$\begin{bmatrix} \cdot405 & \cdot426 & \cdot809 \\ \cdot464 & \cdot076 & -\cdot883 \\ \cdot338 & -\cdot916 & \cdot215 \end{bmatrix} \begin{bmatrix} \cdot405 & \cdot464 & \cdot338 \\ \cdot426 & \cdot076 & -\cdot916 \\ \cdot809 & -\cdot883 & \cdot215 \end{bmatrix} = \begin{bmatrix} 1 & -\cdot494 & -\cdot079 \\ -\cdot494 & 1 & -\cdot103 \\ -\cdot079 & -\cdot103 & 1 \end{bmatrix}$$

This gives the cosines of the angles between the reference vectors and we see that they are obtuse. The angles are approximately:

$$\begin{bmatrix} . & 120^\circ & 95^\circ \\ 120^\circ & . & 96^\circ \\ 95^\circ & 96^\circ & . \end{bmatrix}$$

As soon as we know that the reference vectors are not orthogonal, we have to take account of the fact that the primary factors are not identical with them. Each primary factor is the line in which the hyperplanes intersect, excluding that hyperplane to which the corresponding reference vector is orthogonal. In a three-dimensional common-factor space like ours the primary factors lie along the edges of the pyramid which the extended vectors form.

Let us return to our mental picture, which the reader can place in the room in which he is sitting. The origin, immediately below the point O in Figure 28, is in the middle of the carpet. Figure 28 itself is on the ceiling, seen from above as though translucent. The radial lines with arrowheads are the projections of the primary factors on to the ceiling. The projections of the reference vectors are not drawn, to avoid confusion in the figure. They are near, but not identical with, the primary factors.

The reader should not be misled by the fact that two of the primary factors lie along the same lines as Tests 1 and 7. It was necessary to allow this in devising an example with very few tests in it (to avoid much calculation and printing large tables). But with a large number of tests the lines of the triangle could have been defined without any test being actually at a corner.

3. *Primary factors and reference vectors.*—At about this stage a disturbing thought may have occurred to the reader. We have sought for, and obtained, simple structure on the reference vectors. That is to say, we have found three vectors, three imaginary tests, which are uncorrelated each with a group of the actual tests, namely where there are zeros in the table on page 173. The entries in that table are the projections of the actual tests on the reference vectors.

But the primary factors are different from the reference vectors. The projections of the tests on to the primary factors will be different and will not show these zeros. Those projections are, in fact, given in this table (never mind for the moment how it is arrived at) :

$$F(\Lambda')^{-1}D$$

Structure on the Primary Factors

	L	*B*	*D*
1	·160	·162	·832
2	·408	·666	·793
3	·866	·809	·176
4	·934	·495	·401
5	·541	·927	·152
6	·842	·472	·132
7	·468	·915	·150

Score in Test 1 =	$\cdot 826d$		+ Specific
,, ,, ,, 2 =	$\cdot 536b$	+ $\cdot 701d$	+ ,,
,, ,, ,, 3 =	$\cdot 612l$	+ $\cdot 499b$	+ ,,
,, ,, ,, 4 =	$\cdot 895l$	+ $\cdot 266d$	+ ,,
,, ,, ,, 5 =	$\cdot 880b$		+ ,,
,, ,, ,, 6 =	$\cdot 809l$		+ ,,
,, ,, ,, 7 =	$\cdot 912b$		+ ,,

4. *Behind the scenes.*—It is now time to divulge what these " tests " really are and how the " scores " were made whose correlations we have been analysing, and to compare our analysis with the reality. The example is a simpler and shorter variety of a device used by Thurstone and published in April 1940 in the *Psychological Bulletin.* The measurements behind the correlations were not made on a number of persons, but were made on a number of boxes—only eight boxes, to keep down the amount of calculation and printing. These boxes were of the following dimensions :

	Length	*Breadth*	*Depth*
1	2	2	1
2	3	2	3
3	3	2	2
4	6	3	2
5	4	4	2
6	5	3	1
7	5	4	3
8	4	4	2
Sum	32	24	16
Mean	4	3	2

The " tests " were seven functions of these dimensions, and are shown in the next table, which also shows the score each box (or " person ") would achieve in that test. It is as though someone was unable for some reason to measure the primary length, breadth, and depth of these boxes (as we are unable to measure the primary factors of the mind directly) but was able to measure these more complex quantities like LB, or $\sqrt{(L^2 + D^2)}$ (as we are able to measure scores in complex tests) :

Boxes = Persons

Test	*Formula*	1	2	3	4	5	6	7	8	*Sum*	*Mean*
1	D^2	1	9	4	4	4	1	9	4	36	4·500
2	BD	2	6	4	6	8	3	12	8	49	6·125
3	LB	4	6	6	18	16	15	20	16	101	12·625
4	$\sqrt{(L^2+D^2)}$	2·24	4·24	3·61	6·32	4·47	5·10	5·83	4·47	36·28	4·535
5	$L+B^2$	6	7	7	15	20	14	21	20	110	13·750
6	L^2+D	5	12	11	38	18	26	28	18	156	19·500
7	B	2	2	2	3	4	3	4	4	24	3·000

With these scores the sums of squares and products of deviations from the mean are :

	1	2	3	4	5	6	7
1	66	50·5	22·5	10·2	25	29	3
2	50·5	72·9	98·4	16·8	112·3	100·5	16
3	22·5	98·4	273·9	47·9	259·2	398·5	36
4	10·2	16·8	47·9	11·4	37·0	91·3	4·7
5	25	112·3	259·2	37·0	283·5	288	41
6	29	100·5	398·5	91·3	288	800	36
7	3	16	36	4·7	41	36	6

From these the correlations could be calculated by dividing each row and column by the square root of the diagonal cell entry. But that would make no allowance for specific factors, which in all actual psychological tests play a considerable part. In the example devised by Thurstone on which this is modelled there are no specific factors, but it was decided to introduce them here into Tests 5, 6, and 7, by increasing their sums of squares. In addition, by an arithmetical slip, a small group factor was added to these three tests, and this was not discovered for some time. It was decided to leave it, for in a way it makes the example more realistic, and may be taken to represent an experimental error of some sort running through these three tests.

With these changes, the correlations are found, and are those with which we began this chapter and which we have already analysed into three oblique factors L, B, and D. Let us now compare that analysis with the formulæ which we now know to represent the tests. The pattern on page 177, for example, shows that Test 2 depends only on factors *B* and *D* : and that is correct, for it was, in fact, their product *BD*, and *L* did not enter into it. The analysis gives the test score as a *linear* function of *B* and *D*,

$$\cdot536b + \cdot701d$$

whereas it was really a product. But the analysis was correct in omitting *L*. Similarly, the analyses into the other factors can be compared with the actual formulæ, and in almost every case the factorial analysis, except for being linear, is in agreement with the actual facts. Tests 5 and 6, true, appear in the analysis to omit factors *L* and *D* respectively, although these dimensions figured in their formulæ. But it would appear that they were swamped by reason of the other dimension in the formulæ being squared ; and also possibly the specific and error factors we added did something towards obscuring smaller details. Also the process of " guessing " communalities, though innocuous in a battery of many tests, is a source of considerable inaccuracy when, as here, the tests are few.

5. *Box dimensions as factors.*—We can now explain the *particular* reason for selecting the primary factors, and not the reference vectors, as our fundamental entities. The fundamental entities in the present example can reasonably be said to be the length, breadth, and depth of the boxes, given in the table on page 178. Now, the columns of that table are correlated with one another, as the reader can readily check, the correlation coefficients being—

L with *B*, ·589
L ,, *D*, ·144
B ,, *D*, ·204

These correlations are due to the fact that a long box naturally tends to be large in all its dimensions. It could, of course, be very, very shallow, but usually it is deep and broad.

The reference vectors were, it is true, correlated, but negatively. They were at obtuse angles with one another (see page 174) and obtuse angles have negative cosines corresponding to negative correlations. So the reference vectors do not correspond to the fundamental dimensions length, breadth, and depth.

What, then, are the angles—and hence the correlations—between the primary factors? We shall find that they are acute angles, and their cosines agree reasonably well with the above correlations between the length, breadth, and depth. The algebraic method of finding these angles is given in the mathematical appendix, but it is perhaps desirable to give a less technical account of it here. We need the direction-cosines of the primary factors, that is, the cosines of the angles they make with the orthogonal centroid axes. Each primary factor is the intersection of $n - 1$ hyperplanes—in our simple case is the intersection of two planes.

In n-dimensional geometry a linear equation defines a hyperplane of $n - 1$ dimensions. For example, in a plane of two dimensions a linear equation is a line (of one dimension)—hence the name linear. But in a space of three dimensions a "linear" equation like $ax + by + cz = d$ is a plane. Two such equations define the line which is the intersection of two planes.

Now, the equations of the three planes which form the triangular pyramid of which we have previously spoken are just those equations we have already obtained and used in our example, viz. :

$$\begin{aligned} \cdot405x + \cdot426y + \cdot809z &= 0 \\ \cdot464x + \cdot076y - \cdot883z &= 0 \\ \cdot338x - \cdot916y + \cdot215z &= 0 \end{aligned}$$

These equations taken two at a time define the three edges of the pyramid, which are our primary factors, and if we express each pair in the form—

$$\frac{x}{a} = \frac{y}{b} = \frac{z}{c}$$

then the direction cosines are proportional to a, b, and c, which only require normalizing to be the direction cosines. When the direction cosines are found in this way, and written in columns to form a matrix, they prove to have the values—

$$\begin{bmatrix} \cdot 797 & \cdot 835 & \cdot 503 \\ \cdot 400 & \cdot 187 & -\cdot 843 \\ \cdot 453 & -\cdot 517 & \cdot 192 \end{bmatrix} = (\Lambda')^{-1}D$$

This is the rotating matrix to obtain the projections, i.e. the structure, on the primary factors, and if the centroid loadings on page 172 are post-multiplied by this there results the table we have already quoted on page 175.

The above matrix, premultiplied by its transpose, gives the cosines of the angles between the primary factors. We obtain—

$$\begin{bmatrix} 1 & \cdot 506 & \cdot 150 \\ \cdot 506 & 1 & \cdot 164 \\ \cdot 150 & \cdot 164 & 1 \end{bmatrix} = DC^{-1}D$$

Compare these with the correlations between the columns of dimensions of the boxes, viz. :

$$\begin{bmatrix} 1 & \cdot 589 & \cdot 144 \\ \cdot 589 & 1 & \cdot 204 \\ \cdot 144 & \cdot 204 & 1 \end{bmatrix}$$

The resemblance is quite good, and shows that it is the primary factors, and not the reference vectors, which represent those fundamental although correlated dimensions of length, breadth, and depth in the boxes.

6. *Criticisms of simple structure.*—Thurstone's argument is then, of course, that as this process of analysis leads to fundamental real entities in the case of the boxes (and also in his " trapezium " example, Thurstone, 1944*a*, page 84, with four oblique factors), it may be presumed to give us fundamental entities when it is applied to mental measurements. And I confess that the argument is very strong.

My fears or doubts arise from the possibility that the argument cannot legitimately be reversed in this way. There is no doubt that if artificial test scores are made up with a certain number of common factors, simple structure (oblique if necessary) can be reached and the factors identified. But are there other ways in which the test scores could have been made? Spearman's argument was a similar reversal. If test scores are made with only one common factor, then zero tetrad-differences result. But zero tetrad-differences can be approached as closely as we like by samples of a large number of small factors, with very few indeed common to all the tests.

However, Thurstone's simple structure is a much more complex phenomenon than Spearman's hierarchical order, and yet he seems to have had no great difficulty in finding batteries of tests which give simple structure to a reasonable approximation. I am not sceptical, merely cautious, and admittedly much impressed by Thurstone's ability both in the mathematical treatment and in the devising of experiments.

Thurstone might, I think, put his case in this way. He assembles a battery of tests which to his psychological intuition appear to contain such and such psychological factors, some being memory tests, some numerical, etc., etc., no test, however, containing (to his mind) all these expected factors. He then submits their correlations to his calculations, reaches oblique simple structure, and compares this analysis with his psychological expectation. If there is agreement, he feels confirmed both in his psychology and in the efficacy of his method of finding factors mathematically. Usually there will not be complete agreement, and he is led to modify his psychological ideas somewhat, in a certain direction. To test the truth of these further ideas he again makes and analyses a battery. Especially he looks to see if the same factors turn up in various batteries. He uses his analyses as guides to modifications of his psychological hypotheses, or as confirmation of them. In Great Britain Thurstone's hypothesis of simple structure has been, I think it is correct to say, rather ignored than criticized. Most British psychologists have imbibed during their education a belief in and a partiality for "Spearman's *g*," a factor apparently abolished by Thurstone. Since his work on second-order factors rehabilitates *g*, this objection may disappear. Reyburn and Taylor of South Africa have, however, criticized simple structure shrewdly (1943*a*, and a later paper by Reyburn and Raath, 1949) even although they themselves do not insist on a *g* (see 1941*a*, pages 253, 254, 258).

An early form of response to Thurstone's work was to show that his batteries could also be analysed after Spearman's fashion. Holzinger and Harman (1938), using the Bifactor method, reanalysed the data of Thurstone's *Primary Mental Abilities* and found an important general factor due, as they truly say, " to our hypothesis of its existence and the essentially positive correlations throughout." Spearman (1939*a*) in a paper entitled *Thurstone's Work Reworked* reached much the same analysis, and raised certain practical or experimental objections, claiming that his *g* had merely been submerged in a sea of error. But there is more in it than that. As I said in my contribution to the Reading University Symposium (1939) Thurstone could correct all the blemishes pointed out by Spearman and would still be able to attain simple structure. I said on that occasion that however juries in America and in Britain might differ at present, the larger jury of the future would decide by noting whether Spearman's or Thurstone's system had proved most useful in the hands of the practising psychologist. I now think that they will certainly also consider which set of factors has proved most invariant and most real. Very likely the two criteria may lead to the same verdict. But for the present the two rival claims are in the position described by the Scottish legal phrase, " taken *ad avizandum*."

7. *Application of multiple-factor analysis to industrial test data.*—Dr. R. Harper, with various co-workers, has applied these methods of factor analysis, begun in connexion with psychological tests, to tests of a physical kind on various substances during their manufacture. In *Nature* of November 20th, 1948, Harper and Baron wrote : " In industrial physics there are occasions when empirical tests are employed the exact meaning of which is not fully understood, and where the interrelationships between the tests could profitably be studied by similar means " to those used in psychology, and they described a centroid analysis,

without rotation, of rheological measurements on cheese. In the *British Journal of Applied Physics* of January, 1950, Harper, Kent, and Blair gave an account of the factorial analysis of ten tests (seven rheological and three electrical) on a group of plastics (polyvinyl-chloride-plasticizer mixes). They made a centroid analysis, with four iterations, took out three factors, and rotated them orthogonally to maximize the number of near-zero loadings. They tried also other rotations, including one to an approximate oblique simple structure, and suggest interpretations of the factors arrived at.

CHAPTER XIII

SECOND-ORDER FACTORS

1. *A second-order general factor.*—The reason why the factors arrived at in the " box " example were correlated was that large boxes tend to have all their dimensions large. There is a typical shape for a box, often departed from, yet seldom to an extreme degree. Therefore the length, breadth, and depth of a series of boxes are correlated, and so also are Thurstone's primary factors in such a case. There is a size factor in boxes, a general factor which does not appear as a first-order factor (those we have been dealing with) in Thurstone's analysis, but causes these primary factors to be correlated. Possibly, therefore, when oblique factors appear in the factorial analysis of psychological tests, there is a hidden general factor causing the obliquity. This factor or factors (for there might be more than one) can be arrived at by analysing the first-order factors, into what Thurstone calls second-order factors, factors of the factors.

Of course, whether such a procedure could be justified by the reliability of the original experimental data is very doubtful in most psychological experiments. The superstructure of theory and calculation raised upon those data is already, many would urge, perhaps rather top-heavy, and to add a second storey unwise. But we should not, I think, let this practical question deter us from examining what is undoubtedly a very interesting and illuminating suggestion, which may turn out to be the means of reconciling and integrating various theories of the structure of the mind.

If we take the primary factors of our " box " example of Chapter XII, they were correlated as shown in this matrix :

$$\begin{bmatrix} 1 & \cdot 506 & \cdot 150 \\ \cdot 506 & 1 & \cdot 164 \\ \cdot 150 & \cdot 164 & 1 \end{bmatrix}$$

If we analyse these in their turn into a general factor and specifics we obtain, using the formula—

$$g \text{ saturation} = \left(\frac{r_{ab}r_{ac}}{r_{bc}}\right)^{\frac{1}{2}},$$

the saturations of the primary factors with a second-order g as ·680, ·744, and ·220 ; and each primary factor will also have a factor specific. We have now replaced the analysis of the original tests into three oblique factors by an analysis into four orthogonal factors, one of them general to the oblique factors and presumably also general to the original tests, though that we have still to inquire into. We must also inquire into the relationship of the specifics of the original tests to these second-order factors, which are no longer in the original three-dimensional common-factor space, but in a new space of four dimensions. Are the original test-specifics orthogonal to this new space ?

With only three oblique factors, an analysis into one g is always possible (except in the Heywood case, which will often occur among oblique factors). If there had been four or more oblique factors, we would have had to use more second-order general factors unless the tetrad-differences were zero. Thurstone's " trapezium " example already referred to had four oblique factors, and his article should be consulted by the interested.

2. *Its correlations with the tests.*—Let us turn now to the question what the correlations are between the seven original tests and the above second-order g. To obtain these Thurstone uses an argument equivalent to the following :

We may first note that each reference vector makes an acute angle with its own primary factor, but is at right angles to every other primary factor, for these are all contained in the hyperplane to which it is orthogonal. The cosines of the angles can be obtained by premultiplying the rotation matrix of the reference vectors by the transpose of the rotation matrix of the primary factors.

Correlations between Primary Factors and Reference Vectors

$$D\Lambda^{-1} \times \Lambda = D$$

$$\begin{bmatrix} \cdot797 & \cdot400 & \cdot453 \\ \cdot835 & \cdot187 & -\cdot517 \\ \cdot503 & -\cdot843 & \cdot192 \end{bmatrix} \begin{bmatrix} \cdot405 & \cdot464 & \cdot338 \\ \cdot426 & \cdot076 & -\cdot916 \\ \cdot809 & -\cdot883 & \cdot215 \end{bmatrix} = \begin{bmatrix} \cdot860 & . & . \\ . & \cdot858 & . \\ . & . & \cdot983 \end{bmatrix}$$

These cosines in the diagonal of the matrix D give us the angles 31°, 31°, and 11° which we have already mentioned on page 177 as the angles between each primary factor and its own reference vector.

Each row of the first of the above matrices represents the projections of the primary factor on to the orthogonal centroid axes. These are, in fact, the loadings of the primary factors, thought of as imaginary or possible tests, in the orthogonal centroid factors I, II, and III. Following Thurstone, we add these three rows below the seven rows of our original seven real tests, extending the matrix F in length thus :

$$\begin{bmatrix} F \\ T \end{bmatrix}$$

	I	*II*	*III*	r_g	
1	·449	—·682	·165	·211	wanted
2	·825	—·478	—·129	·574	
3	·906	·336	·020	·787	
4	·846	·133	·457	·666	
5	·808	·208	—·412	·719	
6	·697	·336	·335	·597	
7	·767	·173	—·468	·683	
L	·797	·400	·453	·680	known
B	·835	·187	—·517	·744	
D	·503	—·843	·192	·220	

This lengthened matrix we want to post-multiply by a column vector (ψ in Thurstone's notation) to give the correlations of the tests, including the imaginary tests *L*, *B*, and *D*, with the second-order *g*. In other words, we want to know by what weights each column must be multiplied so that the weighted sum of each row is the correlation of that test with g. Suppose these weights are u, v, and w. Since we already know from our second-order analysis what r_g is for each of the primaries L, B, and D, we have three equations for u, v, and w, the solution of which gives us their values. We have—

$$\cdot 797u + \cdot 400v + \cdot 453w = \cdot 680$$
$$\cdot 835u + \cdot 187v - \cdot 517w = \cdot 744$$
$$\cdot 503u - \cdot 843v + \cdot 192w = \cdot 220$$

and these equations can be solved in the usual way, if the reader wishes. The values are ·798, ·198, and —·077.

A closer examination of them, however, which can be most readily expressed in matrix notation, leads to an easier plan—especially desirable if the number of primary factors were greater. In matrix form the above equations are—

$$\mathrm{T}\psi = r_g$$
$$\text{whence} \quad \psi = \mathrm{T}^{-1} r_g$$

and since T is merely a short notation for $\mathrm{D}\Lambda^{-1}$ we have—

$$\psi = (\mathrm{D}\Lambda^{-1})^{-1} r_g$$
$$= \Lambda \mathrm{D}^{-1} r_g$$

That is to say, the centroid loadings F of the seven tests have to be post-multiplied by this, giving a matrix (a single column)—

$$\mathrm{F}\psi = \mathrm{F}\Lambda \mathrm{D}^{-1} r_g$$

But $\mathrm{F}\Lambda$ we already know. It is (see page 173) the simple structure V on the reference vectors. So we merely have to multiply the columns of V by $\mathrm{D}^{-1} r_g$ and add the rows to get the correlation of each test with g. These multipliers are, that is to say :

$$\cdot 680 \div \cdot 860 = \cdot 791$$
$$\cdot 744 \div \cdot 858 = \cdot 867$$
$$\cdot 220 \div \cdot 983 = \cdot 224$$

The results are the same as by the former method, except for discrepancies due to rounding off decimals, and are given to the right of the preceding table.

3. *A* g *plus an orthogonal simple structure.*—In his own examples, Thurstone has not calculated the loadings of the original tests with the other orthogonal second-order

factors, the factor specifics. This can, however, clearly be done by the same method as above. Since the correlations of the general factor with the three oblique factors are ·680, ·744, and ·220, the correlations of each factor specific with its own oblique factor are ·733, ·668, and ·975. For example, $\cdot733^2 = 1 - \cdot680^2$. The second-order analysis therefore is :

$$\begin{bmatrix} \cdot680 & \cdot733 & . & . \\ \cdot744 & . & \cdot668 & . \\ \cdot220 & . & . & \cdot975 \end{bmatrix} = E$$

Dividing the rows by the divisors already mentioned, viz. ·860, ·858, and ·983, we obtain the matrix :

$$\begin{bmatrix} \cdot791 & \cdot853 & . & . \\ \cdot867 & . & \cdot779 & . \\ \cdot224 & . & . & \cdot992 \end{bmatrix} = D^{-1}E$$

and when the matrix V is post-multiplied by this we obtain the following analysis of the original seven tests into a general factor plus an orthogonal simple structure of three factors :

General Factor plus Simple Structure

$$G = VD^{-1}E$$

	g	λ	β	δ
1	·211	·021	·009	·805
2	·574	·022	·358	·683
3	·787	·449	·333	—·006
4	·666	·656	—·001	·260
5	·719	·071	·588	—·006
6	·597	·593	·041	·000
7	·683	·005	·609	·000

The zero or very small entries in λ, β, and δ are in the same places as they are for L', B', and D' in the oblique simple structure V (see page 173). What we have now done is to analyse the box data into four orthogonal factors corresponding to size, and ratios of length, breadth, and depth. In terms of our pyramidal geometrical analogy we have " taken out a general factor" by depressing the ceiling of our room, squashing the pyramid down until its three plane sides are at right angles to each other.

The above structure, being on orthogonal factors, is also a pattern, so that the inner products of its rows ought to give the correlation coefficients with the same accuracy, if we have kept enough decimal places in our calculations, as do the rows of the centroid analysis F : and so they do. For example, the correlation between Tests 1 and 2 is, from F,

$$\cdot449 \times \cdot825 + \cdot682 \times \cdot478 - \cdot165 \times \cdot129 = \cdot675$$

and from G it is—

$$\cdot211 \times \cdot574 + \cdot021 \times \cdot022 + \cdot009 \times \cdot358 + \cdot805 \times \cdot683 = \cdot675$$

The " experimental " value was ·728, the difference of ·053 being due to the inaccuracy of the guessed communalities, or in an actual experimental set of data to sampling error and to the rank of the matrix not being exactly three.

We can see here a distinct step towards a reconciliation between the analyses of the Spearman school and those of Thurstone using oblique factors. But we must not forget that if the oblique factors are not oblique enough, the Heywood embarrassment will occur, and a second-order *g* be impossible. The orthogonal factors of G are more convenient to work with statistically, but it is possible that the oblique factors of V are more realistic both in our artificial box example and in psychology. They corresponded in our case to the actual length, breadth, and depth of the boxes. The factors λ, β, and δ of matrix G correspond to these dimensions after the boxes have all been equalized in " size."

PART IV

THE ESTIMATION OF FACTORS *

* This use of the word "estimation" has been criticized. By statisticians the word is restricted to mean the estimation of unknown parameters from a sample, a process of inference from sample to parent population. Here the word is used to mean the "estimation" of a man's scores in a test (or vocation or examination) to which he has not been subjected, from a knowledge of his behaviour in other tests. Factors are imaginary tests and a man's score in them can be "estimated" in the same way. I would use another word if I could, but "estimation" seems the natural expression. Besides, I think the two meanings are fundamentally alike.

CHAPTER XIV

REGRESSION AND MULTIPLE CORRELATION

1. *Correlation coefficient as estimation coefficient.*—A correlation coefficient indicates the degree of resemblance between two lists of marks : and therefore it also indicates the confidence with which we can estimate a man's position in one such list x if we know his position in the other y. If the correlation between two lists is perfect ($r = 1$), we know that his standardized score* in the one list is exactly the same as in the other ($x = y$).

If the correlation between the two lists is zero ($r = 0$), then the knowledge of a man's position in the one list tells us nothing whatever about his position in the other list. If we are *compelled* to make an estimate of that, we can only fall back on our knowledge that most men are near the average and few men are very good or very bad in any quality. We have, therefore, most chance of being correct if we guess that this man is average in the unknown test. ($x = 0$. The average mark we have agreed to call zero ; marks above average, positive ; marks below average, negative.)

In the first case, when $r = 1$, we are justified in equating his unknown score x to his known score y—

$$x = y$$

In the second case, when $r = 0$, we are compelled by our ignorance to take refuge in—

$$x = 0 \text{ or average.}$$

Both these statements can be summed up in the one statement—

$$\hat{x} = ry$$

where the circumflex mark over the x is meant to indicate that this is an estimated, not a measured, value. If, now,

* A test score in what follows always means a standardized score unless the contrary is stated. But *estimates* are not in standard measure in general.

we consider a case between these, where the correlation is neither perfect nor zero, it can be shown that this equation still holds, provided each score is measured in standard deviation units. Since r is always a fraction, this means that we always estimate his unknown x score as being nearer the average than his known y score. That is because we know that men tend to be average men. If this man's y score is high, say—

$$y = 2$$

(two standard deviations above the average), and if the correlation between the qualities x and y is known to be $r = \cdot 5$, we guess his position in the x test as being—

$$\hat{x} = ry = \cdot 5 \times 2 = 1$$

i.e. only one standard deviation above the average. This is a guess influenced by our two pieces of knowledge, (1) that he did very well in Test y, which is correlated with Test x, and (2) that most men get round about an average score (zero). It is a compromise, an estimate. It will often be wrong; indeed, very seldom will it be exactly right. But it will be right *on the average*, it will as often be an underestimate as an overestimate, in each array of men who are alike in y. The correlation coefficient, then, is an estimation coefficient for tests measured in standard deviation units.

2. *Three tests.*—Suppose now that we have *three* tests whose intercorrelations are known, and that a man's scores on two of them, y and z, are known. We wish to estimate what his score will most probably be in the other test, x. x need not be a test in the ordinary sense of the word, but may be an occupation for which the man is a candidate or entrant. According as we use his known y or his known z score, we shall have two estimates for his x score. To fix our ideas, let us take definite values for the correlations, say :

	x	y	z
x	1·0	·7	·5
y	·7	1·0	·3
z	·5	·3	1·0

The two estimates for his x are then—

$$\hat{x} = \cdot 7y$$
$$\hat{x} = \cdot 5z$$

and of these we shall have rather more confidence in the estimate associated with the higher correlation. But we ought to have still more confidence in an estimate derived from both y and z. Such an estimate could use not only the knowledge that y and z are correlated with x, but also the knowledge that they are correlated to an extent of $r = \cdot 3$ with each other. Just to take the average of the above two separate estimates will not utilize this knowledge, nor will it utilize the fact that the estimate from y ($r = \cdot 7$) is more worthy of confidence than the estimate from z ($r = \cdot 5$).

What we want is to know how to combine the two scores y and z into a weighted total—

$$(by + cz)$$

which will have the highest possible correlation with x. Such a correlation of a *best*-weighted total with another test is called a *multiple* correlation. From such a weighted total of his two known scores we could then estimate the man's x score more accurately than from either the y or the z score alone. It must use all the information we have, including our information that y and z correlate to an amount $r = \cdot 3$.

3. *The straight sum and the pooling square.*—In order to answer this question, we shall first consider the problem of finding the correlation of the straight unweighted sum of the scores $y + z$ with x. This is the simplest form of a problem to which a general answer was given by Professor Spearman (Spearman, 1913).

We shall put his formula into a very simple form, which we may call a *pooling square*. In our present instance we want to find the correlation of $y + z$ with x (all of these being, we are assuming, measured in standard deviation units). We divide the matrix of correlations by lines separating the " criterion " x from the " battery " $y + z$ thus :

	x	y	z
x	1·0	·7	·5
y	·7	1·0	·3
z	·5	·3	1·0

In each of the quadrants of this pooling square (with unities in the diagonal, be it noted) we are going to form the sum of all the numbers, and we shall indicate these sums by the letters :

A	C
C	B

(where C is the sum of the *C*ross-correlations between the battery $y + z$ and the criterion x, which can be regarded as a second battery of one test only).

Then the correlation of x with $y + z$ is equal to—

$$\frac{C}{\sqrt{AB}}$$

which in our present example is—

$$\frac{\cdot 7 + \cdot 5}{\sqrt{(1) \times (1 + \cdot 3 + \cdot 3 + 1)}} = \frac{1 \cdot 2}{\sqrt{2 \cdot 6}} = \cdot 744$$

so that the battery $(y + z)$ has a rather better correlation (·744) with x than has either of its members (·7 and ·5). From the straight sum of the man's scores in the two tests y and z we can therefore in this case get a better estimate of his score in x than we could get from either alone.

4. *The pooling square with weights.*—We want, however, to know whether a *weighted* sum of y and z will give a still higher combined correlation with x. With sufficient patience, we could answer this by trial and error, for the pooling square enables us to find almost as easily the correlation of a weighted battery with the criterion.* Let us, for example, try the battery $3y + z$. For this purpose

* The pooling square can also be used to find the correlations or covariances of weighted batteries with one another. Elegant developments are Hotelling's ideas of the most predictable criterion (1935*a*) and of vector correlation (1936).

we write the weights along both margins of the pooling square :

		3	1
	1·0	·7	·5
3	·7	1·0	·3
1	·5	·3	1·0

and multiply *both the rows and the columns* by these weights before forming the sums *A*, *B*, and *C*. The result of the multiplications in our case is :

1·0	2·1	·5
2·1	9·0	·9
·5	·9	1·0

=

1·0	2·6
2·6	11·8

and we therefore have—

$$\text{correlation} = \frac{2·6}{\sqrt{11·8}} = ·757$$

a higher value than ·744 given by the simple sum. So we have improved our estimation of the man's x score, and estimates made by taking $3y + z$ would correlate ·757 with the measured values of x.

5. *Regression coefficients and multiple correlation.*—Similarly we could try other weights for y and z and search by trial and error for the best. There is, however, a general answer to this question, namely that *the best weights for y and z are proportional to certain minor determinants of the correlation matrix.* The weight for y is proportional to the minor left when we cross out the criterion column and the y row, the weight for z is proportional to *minus* the minor left when we similarly cross out the criterion column and the z row. The matrix of correlations with the criterion column deleted being :

$$\begin{bmatrix} ·7 & ·5 \\ 1·0 & ·3 \\ ·3 & 1·0 \end{bmatrix}$$

the weight for y is therefore proportional to :

$$\begin{vmatrix} \cdot 7 & \cdot 5 \\ \cdot 3 & 1{\cdot}0 \end{vmatrix} = \cdot 55$$

and that for z is proportional to :

$$-\begin{vmatrix} \cdot 7 & \cdot 5 \\ 1{\cdot}0 & \cdot 3 \end{vmatrix} = \cdot 29$$

that is, they are as ·55 : ·29. To make these weights not merely proportional but absolute values we must divide each of them by the minor left when the row and column concerned with the " criterion " x are deleted, namely :

$$\begin{vmatrix} 1{\cdot}0 & \cdot 3 \\ \cdot 3 & 1{\cdot}0 \end{vmatrix} = \cdot 91$$

so that these absolute best weights, for which the technical name is " regression coefficients," are—

$$\frac{\cdot 55}{\cdot 91}y + \frac{\cdot 29}{\cdot 91}z$$

or

$$\cdot 6044y + \cdot 3187z$$

We are inviting the reader to take this method of calculating the regression coefficients on trust ; but he can at least satisfy himself that when applied to the pooling square they give a higher correlation of battery with criterion than any other weights do. The result of multiplying the y column and row by ·6044, and the z column and row by ·3187, is the following :

		·6044	·3187				
	1·0	·7	·5		1·0000	·4231	·1593
·6044	·7	1·0	·3	=	·4231	·3653	·0578
·3187	·5	·3	1·0		·1593	·0578	·1015

=	1·0000	·5824
	·5824	·5824

$$\text{Multiple correlation} = \frac{\cdot 5824}{\sqrt{\cdot 5824}} = \cdot 763 \doteq r_m, \text{ say, which}$$

is higher than any other weighting will produce, if the reader cares to try others. Notice the peculiarity of the pooling square with regression coefficients as weights, that $C = B$

(·5824 = ·5824). We can deduce that the inner product of the regression coefficients with the correlation coefficients gives the square of the multiple correlation—

$$·604 \times ·7 + ·319 \times ·5 = ·583 = r_m{}^2$$

Indeed, we can take this as forming one reason for using ·604 and ·319, and not any other numbers proportional to them, although the latter would give the same order of merit. We want our estimates of x not merely to be as highly correlated with the true values of x as is possible, but also to be equal to them on the average in the long run, in the sense that our overestimations will, in each array of men who have the same y and z, be as numerous as our underestimations, and this is achieved by using not merely ·55 and ·29 as weights, but ·55 ÷ ·91, and ·29 ÷ ·91.

6. *Aitken's method of pivotal condensation.*—When there are more than two tests y and z in the battery, the application of the above rules becomes increasingly laborious. It is desirable, therefore, to have a routine method of calculating regression coefficients which will give the result as easily as possible even in the case of a team of many tests. The method we shall adopt (Aitken, 1937*a*) is based upon the calculation of tetrads, as already used in our Chapter V. We shall first calculate the above regression coefficients again by this method. Delete the criterion column in the matrix of correlations, *transfer the criterion row to the bottom,* and write the resulting oblong matrix in the top left-hand corner of the sheet of calculations, preferably on paper ruled in squares :

					Check Column
A	(1·0)	·3	−1	.	·3
	·3	1·0	.	−1	·3
	·7	·5	.	.	1·2
B		(·91)	·3	−1	·21
		1·00	·3297	−1·0989	·2308
		·29	·7	.	·99
C			·604	·319	·923

On the right of the oblong matrix of correlation coefficients we rule a middle block of columns of the same number, here *two*, and on the right of all a check column. The columns of the middle block we fill with a pattern of *minus ones* diagonally as shown, leaving the other cells empty,* including the bottom row. In the check column we write the sum of each row. The top left-hand number of all we mark as the "pivot." Slab *B* of the calculation is then formed from slab *A* by writing down, in order as they come, all the tetrad-differences of which the pivot in *A* is one corner. Thus the first row of slab *B* is calculated thus—

$$\begin{aligned} 1 \times 1 - \cdot 3 \times \cdot 3 &= \cdot 91 \\ 1 \times 0 - \cdot 3 \times (-1) &= \cdot 3 \\ 1 \times (-1) - \cdot 3 \times 0 &= -1 \\ 1 \times \cdot 3 - \cdot 3 \times \cdot 3 &= \cdot 21 \end{aligned}$$

and the row is checked by noting that ·21 is the sum of the others. Immediately below this first row a second version of it is written, with every member divided by the first (·91). This is to facilitate the calculation of slab *C* by having unity again as a pivot. The second row of slab *B* is then formed, beginning with—

$$1 \times \cdot 5 - \cdot 7 \times \cdot 3 = \cdot 29$$

Throughout the whole calculation, except for the division of the first row, only one operation needs to be performed, namely the computing of tetrad-differences, beginning with the pivot.

The same operation is then repeated to give slab *C*, using the modified first row of *B*, with pivot unity.

This procedure goes on, slab after slab, until no numbers remain in the left-hand block. There being only three tests in all in our example, this happens at slab *C*. The middle block then gives the regression coefficients ·604 and ·319, with their proper signs, all ready for use. Throughout the calculation the check column detects any blunder in each row. The check, let me repeat, for I often find this misunderstood, consists in seeing that the appropriate

* The dots represent zeros.

tetrad from the sums in the previous slab agrees with the sum of the new row. Thus ·99 is both the sum of its row, and also the tetrad—

$$1 \times 1{\cdot}2 - {\cdot}7 \times {\cdot}3$$

from slab *A*.

When the number of tests in the battery is large, the calculation of the regression coefficients is a laborious business, but probably less so by this method than by any other. It will be clear to the reader that so long a calculation is not worth performing unless the accuracy of the original correlation coefficients is high. Only very accurate values can stand such repeated multiplication, etc., without giving untrustworthy results. In other words, regression coefficients have a rather high standard error.*

7. *A larger example.*—Next we give in full the calculation of the regression coefficients in a slightly larger example, though one still much smaller than a practical scheme of vocational advice would involve. Here z_0 is the " occupation," and z_1, z_2, z_3, and z_4 are tests. To give the example an air of reality, these and their intercorrelations are taken from Dr. W. P. Alexander's experimental study, *Intelligence, Concrete and Abstract* (Alexander, 1935). They were †:

- z_1 Stanford-Binet test ;
- z_2 A picture-completion test ;
- z_3 Thorndike reading test ;
- z_4 Spearman's analogies test in geometrical figures.

* Regression weights obtained from one set of data, applied to a subsequent set, will not usually give a correlation with the criterion as high as that predicted. The probable defect in its square will be (Wherry, 1931)—

$$(1 - r^2_m)(M - 1)/(N - M),$$

where N is the number of persons and M the number of tests.

† In this, as in other instances where data for small examples are taken from experimental papers, neither criticism nor comment is in any way intended. Illustrations are restricted to few tests for economy of space and clearness of exposition, but in the experiments from which the data are taken many more tests are employed, and the purpose may be quite different from that of this book.

But the occupation is a pure invention, for purposes of this illustration only. The correlation matrix is :

	z_0	z_1	z_2	z_3	z_4
z_0	1.00	·72	·63	·58	·41
z_1	·72	1·00	·39	·69	·49
z_2	·63	·39	1·00	·19	·27
z_3	·58	·69	·19	1·00	·38
z_4	·41	·49	·27	·38	1·00

The fact that we possess these correlations means that we have given these tests to a sufficiently large number of persons whose ability in the occupation is also known. The occupation can be looked upon as another test, in which marks can be scored. In an actual experiment, obtaining marks for these persons' abilities in the occupation is in fact one of the most difficult parts of the work. We can now find by Aitken's method the best weights for Tests z_1 to z_4 to make their weighted sum correlate as highly as possible with z_0. For a reason which will be explained later, I have numbered the tests in the order of their correlations with the criterion. To make the arithmetic as easy as possible to follow in an illustration, the original correlation coefficients are given to two places of decimals only, and only three places of decimals are kept at each stage of the calculation. The previous explanation ought to enable the reader to follow. As an additional help, take the explanation of the value ·454 in the middle of slab *C*. It is obtained thus from slab *B*—

$$1 \times \cdot 490 - \cdot 079 \times \cdot 460 = \cdot 454$$

and is typical of all the others. Except for the division of each first row, only one kind of operation is required through the whole calculation, which becomes quite mechanical. The numbers shown on the left in brackets are the reciprocals of ·848, ·517, ·748, used as multipliers instead of dividing by the latter numbers, in obtaining the modified first rows. The process continues until the left-hand block is empty, when the regression coefficients appear in the middle block.*

The result is that we find that the best prediction of a man's probable success in this occupation is given by the regression equation—

$$\hat{z}_0 = \cdot 390 z_1 + \cdot 431 z_2 + \cdot 222 z_3 + \cdot 018 z_4$$

We give a candidate the four tests, reduce his scores

COMPUTATION OF REGRESSION COEFFICIENTS

Aitken's Modified Method with Each Pivot converted to Unity

									Check
A	(1)	·39	·69	·49	−1	.	.	.	1·57
	·39	1	·19	·27	.	−1	.	.	·85
	·69	·19	1	·38	.	.	−1	.	1·26
	·49	·27	·38	1	.	.	.	−1	1·14
	·72	·63	·58	·41	.	.	.	.	2·34
	(1·179)	(·848)	−·079	·079	·390	−1	.	.	·238
B		1·000	−·093	·093	·460	−1·179	.	.	·281
		−·079	·524	·042	·690	.	−1	.	·177
		·079	·042	·760	·490	.	.	−1	·371
		·349	·083	·057	[·720]	.	.	.	1·209
	(1·936)		(·517)	·049	·726	−·093	−1	.	·199
C			1·000	·096	1·406	−·180	−1·936	.	·386
			·049	·753	·454	·093	.	−1	·349
			·116	·025	[·559]	[·412]	.	.	1·112
	(1·337)			(·748)	·384	·102	·095	−1	·329
D				1·000	·514	·136	·128	−1·337	·441
				·014	[·397]	[·433]	[·224]	.	1·068
E					·390	·431	·222	·018	1·061
					Final Regression Coefficients				

* The product of all the unconverted pivots, 1 × ·848 × ·517 × ·748, is the value ·328 of the determinant:

$$\begin{vmatrix} 1\cdot 00 & \cdot 39 & \cdot 69 & \cdot 49 \\ \cdot 39 & 1\cdot 00 & \cdot 19 & \cdot 27 \\ \cdot 69 & \cdot 19 & 1\cdot 00 & \cdot 38 \\ \cdot 49 & \cdot 27 & \cdot 38 & 1\cdot 00 \end{vmatrix}$$

If this alone were wanted, the middle block, and the criterion row, would, of course, be unnecessary.

to standard measure by dividing by the known standard deviation of each test, insert these standard scores into this equation, and obtain an estimated score for him in the occupation. Thus the following three young men could be placed in their probable order of efficiency in this occupation from their test scores :

	Standard Scores in				
	z_1	z_2	z_3	z_4	$\hat{z}_0$
Tom	·7	·0	·2	− ·5	·31
Dick	− ·4	− ·8	·1	·3	− ·47
Harry	·2	1·3	·8	·6	·83

The multiple correlation of such estimates $\hat{z}_0$ with the true values would be obtained by inserting the four correlation coefficients—

$$\cdot 72 \qquad \cdot 63 \qquad \cdot 58 \qquad \cdot 41$$

instead of the z's in the regression equation, and taking the square root, thus—

$$\cdot 390 \times \cdot 72 + \cdot 431 \times \cdot 63 + \cdot 222 \times \cdot 58 + \cdot 018 \times \cdot 41$$
$$= \cdot 68847 = r_m^2$$
$$\therefore r_m = \cdot 83$$

Finally, we can, as we did in the former example, use the regression weights on a pooling square and see if we obtain this same multiple correlation of $r_m = \cdot 83$:

		·390	·431	·222	·018
	1·00	·72	·63	·58	·41
·390	·72	1·00	·39	·69	·49
·431	·63	·39	1·00	·19	·27
·222	·58	·69	·19	1·00	·38
·018	·41	·49	·27	·38	1·00

It will be remembered that we have to multiply each row and column by its appropriate weight, and then sum all the numbers in each quadrant. The easiest way of doing this in large pooling squares is to multiply the rows first, then add the columns and multiply the totals by the column weights, finally adding these products, thus :

Multiply the rows :

		·390	·431	·222	·018
	1·0000	·72	·63	·58	·41
	·2808	·3900	·1521	·2691	·1911
	·2715	·1681	·4310	·0819	·1164
	·1288	·1532	·0422	·2220	·0844
	·0074	·0088	·0049	·0068	·0180
Sums	·6885	·7201	·6302	·5798	·4093

If we had kept all decimals these columnar sums would, since we are using regression coefficients as weights, have been exactly equal to the top row. With the actual figures shown, on multiplying the column totals and adding them, we find that the pooling square condenses to :

1·0000	·6885
·6885	·6885

$$r_m = \frac{\cdot 6885}{\sqrt{\cdot 6885}} = \cdot 83 \text{ as before.}$$

8. *Using fewer tests.*—There is a tendency, which common sense finds natural, for the regression coefficients of the tests of a battery to be in the same order of magnitude as their correlations with the criterion. But this is not invariably the case, and in the present example, if we compare the two sets—

correlations with criterion	·72	·63	·58	·41
and regression coefficients	·390	·431	·222	·018,

we see that Test 2 has a higher regression coefficient than Test 1, although a lower criterion correlation. The reason lies in the high correlation of Test 1 with Test 3, ·69. They measure to that extent the same thing, and when

Test 3 is introduced into the battery it begins to some extent to put Test 1's "nose out of joint."

The boxed numbers in the calculation on page 205 are all regression coefficients. If only Test 1 is used, its regression coefficient is ·72. If Tests 1 and 2 are used, their regression coefficients are ·559 and ·412. If Tests 1, 2, and 3 are used, their regression coefficients are ·397, ·433, and ·224. And if all four Tests are used, the four final numbers are the regression coefficients.

The addition of each test raises the multiple correlation r_{max}. We have—

		r^2_{max}
Test 1	$\cdot72 \times \cdot72$	$= \cdot5184$
Tests 1 and 2	$\cdot72 \times \cdot559 + \cdot63 \times \cdot412$	$= \cdot6622$
Tests 1, 2, and 3	$\cdot72 \times \cdot397 + \cdot63 \times \cdot433 + \cdot58 \times \cdot224$	$= \cdot6882$
Tests 1, 2, 3, and 4	$\cdot72 \times \cdot390 + \cdot63 \times \cdot431 + \cdot58 \times \cdot222 + \cdot41 \times \cdot018$	$= \cdot6885$

Although the addition of each test raises the multiple correlation, some do so only very little ; and our caution in ordering the tests in accordance with the magnitude of the criterion correlation makes it probable, though not certain, that the comparatively useless tests will be the later ones. We can at each stage of the calculation pause and see whether the test we have just added makes a significant addition to the multiple correlation. We do this by an analysis of variance (see Lindquist, 1940, Chapter V, or other text-book). Consider, for example, the rise in the squared multiple correlation from ·6622 to ·6882. Is the rise statistically significant ? To decide this we must know the number of persons tested, say N = 105.

Tests	r^2_{max}	*Degrees of Freedom*	*Mean Square*	*Ratio F*
1 and 2 . .	·6622	2		
Increment on adding 3 . .	·0260	1	·0260	·0260÷·0031=7·7
Residue . .	·3118	101	·0031	
Total . .	1·0000	N−1=104		

The calculation is carried out in the above form, and the decision whether the increment of r^2_{max} is statistically significant depends on the size of the ratio F. If it is large enough, the increase is significant. To decide how large, consult Table V in Fisher and Yates's *Statistical Tables*, where we find that, with degrees of freedom 1 and 101, a ratio of 6·88 would be significant at the 1 per cent. point, i.e. quite highly significant, and 7·7 is even larger than this. So the increase due to the addition of Test 3 is well worth while. A similar calculation for the further addition of Test 4, producing a rise of ·0003 in r^2_{max}, shows, as might be expected, that this is not significant, for F is now less than unity, and Tests 1, 2, and 3 are (with 105 cases) as good as the whole battery.

Tests	r^2_{max}	*Degrees of Freedom*	*Mean Square*	*Ratio F*
1, 2, and 3 . .	·6882	3		
Increment on adding 4 . .	·0003	1	·0003	Less than unity.
Residue . .	·3115	100	·0031	
Total . .	1·0000	104		

9. *Calculation of a reciprocal matrix.*—A somewhat longer method of calculating regression coefficients has two advantages : it permits the easy calculation of regression coefficients for any criterion (or many) when once the main part of the computation is completed, and, what is of great importance, it enables the standard errors of the coefficients, and of their differences, to be found quickly.

The method referred to is to find first of all the reciprocal of the matrix of correlations of the tests. This is done by pivotal condensation also, as illustrated in the table overleaf. The matrix whose reciprocal is required appears in the top left-hand corner, with a diagonal array of *minus ones* on its right, and a diagonal of *plus ones* below it. The whole is condensed in the manner already described on page 205, and the required reciprocal matrix and also that nearly half the numbers can be written down from symmetry.

The regression coefficients for any criterion are then obtained by multiplying the rows of the reciprocal by the criterion correlations and then adding the columns. In the example of page 205 we multiply the first row of the reciprocal by ·72, the second by ·63, and so on. The addition of the columns then gives the same regression coefficients as were found on page 205.

10. *Variances and covariances of regression coefficients.*—The most important advantage of this method is that *whatever the criterion*, the variances and covariances of the regression coefficients are proportional to the cells of the above reciprocal matrix (Fisher, 1925, 15 and 1922, 611). This enables their absolute values for any given criterion to be obtained by multiplying by $1 - r^2_m$ (the defect of the square of the multiple correlation from unity), and dividing by the number of "degrees of freedom" which is for full correlations $N - p - 1$ where N is the number of persons tested, and p the number of tests. For partial correlations the degrees of freedom are reduced by the number of variables "partialled out."

Thus in our example, where $p = 4$, if N had been 105, $N - p - 1$ would be 100. The multiple correlation was ·83, and $1 - r^2_m = {\cdot}312$ (see page 206). The variances and covariances of our four regression coefficients are in this case equal to the reciprocal matrix multiplied by ·00312.

·0075	—·0017	—·0042	—·0016
—·0017	·0038	·0006	—·0004
—·0042	·0006	·0061	—·0004
—·0016	—·0004	—·0004	·0042

The standard errors of the regression coefficients are the square roots of the diagonal elements:

Regression coefficients	·390	·431	·222	·018
Standard errors	·087	·062	·078	·065
Significant ?	Yes	Yes	?	No

The correlations of the regression coefficients will be got by dividing each row and column by the square root of the diagonal element. We obtain :

1·00	—·31	—·62	—·28
—·31	1·00	·12	—·10
—·62	·12	1·00	—·08
—·28	—·10	—·08	1·00

We can now calculate the standard error of the difference between any pair of the regression coefficients and see whether they differ significantly. Take, for example, those for Test 1 (·390) and Test 2 (·431). The difference is ·041. Its standard error is the square root of

$$\cdot 0075 + \cdot 0038 + 2 \times \cdot 31 \times \cdot 087 \times \cdot 062 = \cdot 0146$$

$$\therefore \text{ standard error of } \cdot 041 \text{ is } \cdot 121$$

The difference is therefore not significant when N = 105. Had N been larger it might have been.

11. *The geometrical picture of regression.*—Before we close this chapter it will be illuminating to consider what regression and estimation mean in terms of the geometrical picture of Chapter VI. Consider the illustration used in the earlier pages of the present chapter, with the matrix :

	x	y	z
x	1·0	·7	·5
y	·7	1·0	·3
z	·5	·3	1·0

Here x is the criterion, y and z are the tests. Each of them can be represented by a directed line, as explained in Chapter VI, with angles between these lines such that their cosines are the above correlations. The three lines will then be in an ordinary space of three dimensions.

The *two* tests y and z themselves have, of course, lines which lie in a plane : any two lines springing from the same point as origin lie in a plane. The criterion line x is not in this plane (say, the table top, on which we may imagine lines y and z to lie), but makes an angle with it. The problem of regression and multiple correlation is, in terms of this geometrical picture, to find the line in the plane of y and z which makes the smallest possible angle with the line x : for the smallest possible angle corresponds to the largest possible correlation. Clearly this desired line is the line which is the projection of the line x on to the yz plane, the shadow thrown by x on the table with the sun right overhead. In Figure 29 it is the line OB, where B is vertically below a point A on the test line x.

The regression coefficients are numbers which express the proportions in which the tests y and z have to be combined to give this line OB. It is just like the parallelogram

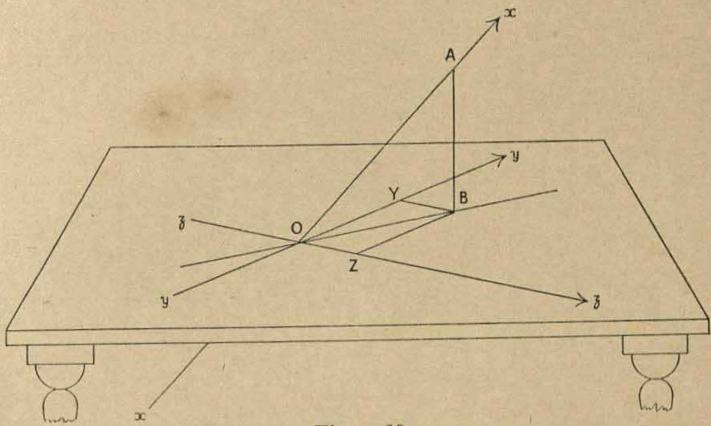


Figure 29.

of forces. If from B we draw parallels to the two test lines, we obtain OY and OZ as the distances to be measured along the two test lines to give a resultant along OB, which is as near as we can come to OA. (No combination of y and z can give a line out of their plane.) *If the distance OA is taken as unity, the distances OY and OZ are the actual regression coefficients.* If a wire model like Figure 29 were made with the proper angles with cosines x with y equal to ·7, x with z equal to ·5, and y with z equal to ·3, the distances OY and OZ would be found to be ·6044 and ·3187. And the cosine of the angle BOA would be ·763, the value we found for the multiple, or highest possible, correlation of the two-test battery with x in Section 5 of this chapter, page 200.

12. *Estimation the same as projection.*—Let us now consider a man P whose two scores in the Tests y and z we know, and whose probable score in Test x we wish to

estimate. His two scores OM and ON in y and z enable us to assign to this man a point P on the yz plane, a point so chosen that its projections on to the y and z vectors give the scores made by him in those tests (see Figure 30). *But we cannot say that this is his point in the three-dimensional space of x, y, and z.* His point in that space may be anywhere on a line $P'PP''$ at right angles to the plane yz. For

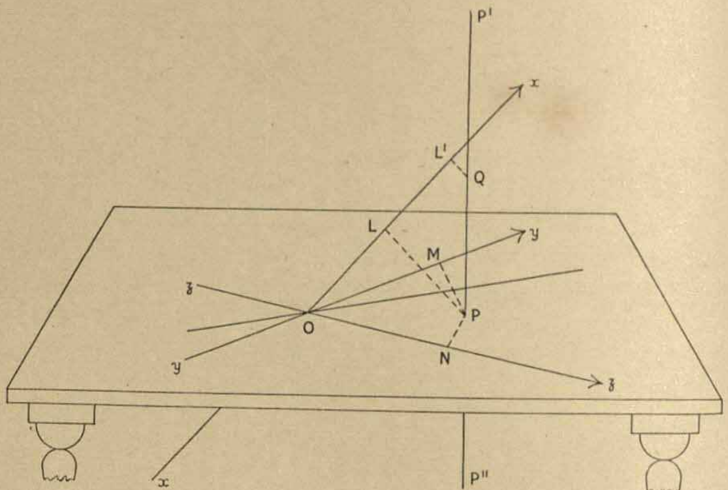


Figure 30.

from anywhere on that line, projections on to y and z fall on the points M and N. Yet the projection on to the vector x, which gives his score in the criterion test x, depends very much on the position of his point on the line $P'PP''$. All the people represented by points on that line have the same scores in y and z but different scores in x, and our man may be any one of them. Before deciding what to do in these circumstances, let us consider this set of people $P'PP''$ in more detail.

It will be remembered that the whole population of persons is represented by a spherical swarm of points, crowded together most closely round about the origin O, and falling off in density equally in all directions from that point. Every test line is a diameter of this sphere, and the plane containing any two test vectors divides the spherical swarm into equal hemispheres. It follows that a line like $P'PP''$ is a chord of the sphere at right angles to a diameter (the line OP), and consequently that it is peopled symmetrically on both sides of P, both upwards along PP' in our figure, and downwards along PP'', the men on the line being most crowded near the point P itself. The average man of the array of men $P'PP''$ (who are all alike in their scores in the two tests y and z) is therefore the man at P, and since we do not know exactly where our candidate's point is along $P'PP''$, we take refuge in guessing that he is the average man of his group and is at the point P itself. From P, therefore, we drop a perpendicular on to the vector x, and take the distance OL as representing his estimated score in that test. This geometrical procedure corresponds exactly to the calculation we made, as a little solid trigonometry will show the mathematical reader. The non-mathematical reader must take it on trust, but the model may illuminate the calculation. OL is the *average* of all the different scores x that a person with scores OM and ON can have. The estimate will only be certain if the line x itself is on the table ; it will be less and less certain, the more the line x is inclined to the table.

It should be noted that the angles which three test vectors make with each other are impossible angles, if the determinant of the matrix of correlations becomes negative. Ordinarily, that determinant is positive. In our present example we have, for example :

$$\begin{vmatrix} 1{\cdot}0 & {\cdot}7 & {\cdot}5 \\ {\cdot}7 & 1{\cdot}0 & {\cdot}3 \\ {\cdot}5 & {\cdot}3 & 1{\cdot}0 \end{vmatrix} = {\cdot}38$$

Such a determinant, however, though it cannot be negative, can be zero, namely in the cases where the two smaller angles exactly equal the largest. In that case the three vectors lie in one plane—the criterion line has sunk until it too lies on the table. In that case alone, when the determinant is zero, the " estimation " is certain, and all the people in the line $P'PP''$ have not only the same scores in y and z, but also the same scores in x. The vanishing of the above determinant therefore shows that this is so. And in more than three dimensions, although we can no longer make a model, the vanishing of the determinant :

$$\begin{vmatrix} 1 & r_{01} & r_{02} & r_{03} & \cdot & r_{0n} \\ r_{01} & 1 & r_{12} & r_{13} & \cdot & r_{1n} \\ r_{02} & r_{12} & 1 & r_{23} & \cdot & r_{2n} \\ r_{03} & r_{13} & r_{23} & 1 & \cdot & r_{3n} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ r_{0n} & r_{1n} & r_{2n} & r_{3n} & \cdot & 1 \end{vmatrix} = \Delta, \text{ say,}$$

shows that the criterion z_0 can be exactly estimated from the team $z_1, z_2 \ldots z_n$. In fact, the multiple correlation r_m, which we have already learned to calculate in another way, can also be calculated as—

$$r_m = \sqrt{1 - \frac{\Delta}{\Delta_{00}}}$$

where Δ is the whole determinant, and Δ_{00} is the minor left after deleting the criterion row and column. This expression clearly becomes equal to unity when $\Delta = 0$. In our small example x, y, z, we have—

$$\Delta = \cdot 38 \qquad \Delta_{00} = \cdot 91$$

$$r_m = \sqrt{1 - \frac{\cdot 38}{\cdot 91}} = \sqrt{\frac{\cdot 53}{\cdot 91}} = \sqrt{\cdot 5824} = \cdot 763$$

as we already know it to be from page 200.

13. *The "centroid" method and the pooling square.*—The pooling square, which we have learned to use in this chapter, enables us to see in another light the nature of the factors first arrived at by the "centroid" method.

Equal Weights

		z_1	z_1	z_2	z_3	z_4
	z_1	1	1	r_{12}	r_{13}	r_{14}
Equal Weights	z_1	1	1	r_{12}	r_{13}	r_{14}
	z_2	r_{12}	r_{12}	1	r_{23}	r_{24}
	z_3	r_{13}	r_{13}	r_{23}	1	r_{34}
	z_4	r_{14}	r_{14}	r_{24}	r_{34}	1

Let us suppose that the tests z_1, z_2, z_3, and z_4 have the correlations shown, and let us by the aid of a pooling square find the correlation of each of them with the average of all. This means giving each test an equal weight in pooling it.

The correlation of z_1 with the average of all is then obtained from the above pooling square (see previous page), which condenses to :

1	$1 + r_{12} + r_{13} + r_{14}$
1 $+ r_{12}$ $+ r_{13}$ $+ r_{14}$	*Sum of all the cells of the table of correlations.*

and the correlation coefficient is—

$$\frac{1 + r_{12} + r_{13} + r_{14}}{\sqrt{\text{above sum}}}$$

This, however, *is* exactly the centroid or simple summation process applied to a table with full communalities of unity. The first centroid factor obtained from such a table is simply for each individual the average of his four test scores, and the method is called the " centroid " method, because " centroid " is the multi-dimensional name for an average (*Vectors*, Chapter III ; and see Kelley, 1935, 59). The line in our geometrical picture, which represents the first centroid factor, is in the midst of the radiating lines which represent the tests, like the stick of a half-opened umbrella among the ribs. It does not, however, make equal angles with the test lines unless these all make equal angles with each other. If several of them are clustered together, and the others spread more widely, the factor will lean nearer to the cluster.

In the foregoing explanation the communalities have been taken as unity, and the factor axis was pictured in the midst of the test lines. If smaller communalities are used, the only difference is that a specific component of each test is discarded, and the first-factor axis must be pictured as in the midst of the lines representing the other components of the tests. It can be shown that when communalities less than unity are used, if we bear in mind that the communal components of the tests are not then standardized, the pooling square gives the correlations exactly as before, if we use communalities instead of units in the diagonal.

The first centroid factor is the average of the communal parts of the tests.

The later factors in their turn are, in a sense, averages of the residues. There are, however, some complications, the first being that the average of the residues just as they stand is zero. The manner in which Thurstone circumvents this has already been described in Chapter V.

14. *The most predictable criterion.*—Often a criterion is also composed of parts, just as a battery of tests is. If it is success in an occupation, the journeyman may be judged for skill, for regularity of attendance, for his manner in dealing with colleagues or customers, etc. Some of these items will consciously or unconsciously be weighted more heavily than others in an adjudicator's assessment of the man ; and so too in the assessment of a boy's success in a secondary school. If the weights are thus decided by employer, or by headmaster, the criterion score becomes again one number, the sum of the arbitrarily weighted parts.

Hotelling, however, raised and solved the question of how to weight the parts of a criterion so that it would correlate most highly with a given battery of tests, also weighted in its best way (Hotelling, 1935*a*, and see Thomson 1947, 1948, and M. S. Bartlett, 1948). There are, then, indeed two weighted batteries. In terms of our geometrical analogy, the criterion is now no longer a line, as in Figures 29 and 30, but a space, and the problem is to find a line in the criterion space, and one in the battery space, which will be as near to each other as possible, both springing from an origin *O* common to both spaces. This technique, which the reader will find illustrated by an arithmetical example in Thomson (1947, 1948), would, for instance, enable weights to be given to the tests in two different batteries to make these agree with one another as much as possible.

15. *Weighting for battery reliability.*—A special case arises when the two batteries are composed of alternative forms of the same tests, when the correlation between the two batteries is the battery reliability, which can be enhanced by suitable weighting.

Thomson (1940) described how to find the best weights for battery reliability, as a special case of Hotelling's " most predictable criterion," and Peel (1947) has given a simpler formula than Thomson's (see page 353 in the Mathematical Appendix, Section 9*a*). If there are only two tests in the battery, with reliabilities r_{11}, r_{22} and correlating with one another r_{12}, then Peel's formula gives as the maximum attainable reliability the largest root μ of the equation.

$$\begin{vmatrix} r_{11} - \mu & r_{12}(1-\mu) \\ r_{12}(1-\mu) & r_{22} - \mu \end{vmatrix} = 0$$

that is $\mu^2(1 - r_{12}^2) - \mu(r_{11} + r_{22} - 2r_{12}^2) + (r_{11}r_{22} - r_{12}^2) = 0$. If, for example, $r_{12} = \cdot 5$, $r_{11} = \cdot 7$, and $r_{22} = \cdot 8$, the quadratic has roots ·843 and ·490, and a battery reliability of ·843 is attainable by using weights proportional to either row of the above determinant with $\mu = \cdot 843$, taken reversed and with alternate signs, that is ·0785 and ·1431

or ·0431 and ·0785

or 1 and 1·8 approximately.

If as a check we set out a pooling square for the two batteries it will be—

	1	1·8	1	1·8
1	1·0	·5	·7	·5
1·8	·5	1·0	·5	·8
1	·7	·5	1·0	·5
1·8	·5	·8	·5	1·0

and if we multiply the rows and columns by the weights shown, and add together the quadrants, this reduces to—

6·04	5·092
5·092	6·04

giving a battery self-correlation or reliability of—

$$\frac{5{\cdot}092}{6{\cdot}04} = {\cdot}843 \text{ as expected.}$$

When there are more than two tests, the solution of the above determinantal equation becomes laborious and difficult. Green (1950) has given a transformation of the equation which enables an iterative process to be used in its solution, making it more practicable (see the Mathematical Appendix, page 353).

Clearly the weights making a battery as reliable as possible will not be the same as those making it most valid in predicting a given criterion. There is here a conflict of aims, for we want a battery to be both as valid and as reliable as possible. It is very desirable that some reasonably simple form of calculation should be devised to find those weights which should be given to the tests of a battery which, for a given criterion, would make the best compromise, making reliability equal to validity and both as great as possible (see Thomson 1940, pages 364 to 365).

CHAPTER XV

THE ESTIMATION OF A MAN'S FACTORS

1. *Estimating a man's "g."*—So far, our discussion of estimation in Chapter XIV has had nothing immediate to do with factorial analysis. We are next, however, going to apply these principles of estimation to the problem of estimating a man's factors, given his test scores. As we have already explained in Chapter VII, there is no need to "estimate" factors when unity is retained in each diagonal cell; they can be calculated without any loss of exactness because they are equal in number to the tests: and even if we analyse out only a few of them, they can be exactly calculated for a man from his test scores. When we say *exactly* here, we mean that the factors are known with the same exactness as the test scores which are our data.

When communalities are used, however, factors are more numerous than the tests, and can therefore only be "estimated." Two men with the same set of test scores may have different factors. All we can do is to estimate them, and since the test scores of the two men are the same, our estimates of their most probable factors will be the same. The problem does not differ essentially from the estimation of occupational success or of ability in any "criterion" test. The loadings of a factor in each test give the z_0 row and column of the correlation matrix. Let us first consider the case of a hierarchical battery of tests, and the estimation of g, taking for our example the first four tests of the Spearman battery used as illustration in Chapter I, with these correlations:

	z_1	z_2	z_3	z_4
z_1	1·00	·72	·63	·54
z_2	·72	1·00	·56	·48
z_3	·63	·56	1·00	·42
z_4	·54	·48	·42	1·00

These correspond, in the analogy with the ordinary cases of estimation of the first chapter of this part, to the tests given to a candidate. In those cases, however, there was a real criterion whose correlations with the team of tests were known, and formed the z_0 row and column of the matrix. Here the "criterion" is g, and it cannot be measured directly; it can only be estimated in the manner we are now about to describe. We have here, therefore, no row and column of experimentally measured correlations for the criterion z_0 or g in the present case (Thomson, *B.J.P.* **25**, 94). From the hierarchical matrix of inter-correlations of the tests, however, we can calculate the "saturation" or "loading" of each test with the hypothetical g, and use these for our criterion column and row of correlations. These saturations are the correlation coefficients which would be found between each test and a test of pure g with no specific. We thus arrive at the matrix:

	z_0	z_1	z_2	z_3	z_4
z_0	1·00	·90	·80	·70	·60
z_1	·90	1·00	·72	·63	·54
z_2	·80	·72	1·00	·56	·48
z_3	·70	·63	·56	1·00	·42
z_4	·60	·54	·48	·42	1·00

and we want to know the best-weighted combination of the test scores z_1 to z_4 in order to correlate most highly with $z_0 = g$. The problem is now the same as one of ordinary estimation of ability in an occupation, and the mathematical answer is the same. We can, for example, use Aitken's method of finding the regression coefficients, although in this case, because of the hierarchical qualities of the matrix, there is, as we shall shortly see, an easier method. It is, however, illuminating for the student actually to work out the regression coefficients as in an ordinary case of estimation, as shown on the next page.

If, therefore, we know the scores z_1, z_2, z_3, and z_4 which a man has made in these four tests, we can estimate his g by the equation—

$$\hat{g} = \cdot5531z_1 + \cdot2595z_2 + \cdot1602z_3 + \cdot1095z_4$$

(1·00)	·72	·63	·54	−1·00	.	.	.	1·89
·72	1·00	·56	·48	.	−1·00	.	.	1·76
·63	·56	1·00	·42	.	.	−1·00	.	1·61
·54	·48	·42	1·00	.	.	.	−1·00	1·44
·90	·80	·70	·60	.	.	.	.	3·00
(2·0764)	(·4816)	·1064	·0912	·72	−1·00	.	.	·3992
	1·0000	·2209	·1894	1·495	−2·0764	.	.	·8289
	·1064	·6031	·0798	·63	.	−1·00	.	·4193
	·0912	·0798	·7084	·54	.	.	−1·00	·4190
	·1520	·1330	·1140	·90	.	.	.	1·2994
(1·7253)		(·5796)	·0596	·4709	·2209	−1·00	.	·3311
		1·0000	·1028	·8124	·3811	−1·7253	.	·5712
		·0597	·6911	·4037	·1894	.	−1·00	·3438
		·0994	·0852	·6728	·3156	.	.	1·1730
(1·4599)			(·6850)	·3552	·1666	·1030	−1·00	·3097
			1·0000	·5186	·2432	·1504	−1·4599	·4521
			·0750	·5920	·2777	·1715	.	1·1162
				·5531	·2595	·1602	·1095	1·0823

Regression Coefficients

The multiple correlation of such estimates in a large number of cases with the true values of g will be by analogy with our former case given by—

$$r_m^2 = ·5531 \times ·90 + ·2595 \times ·80 + ·1602 \times ·70 + ·1095 \times ·60 = ·883$$

$$r_m = ·940$$

We must remember, however, that such a correlation here is rather a fiction. We had in the former case the possibility of comparing our estimates with the candidate's eventual performance in the occupation or criterion z_0. Here we have no way of knowing g; we only have the estimates.

As before, we can check the whole calculation by a pooling square (see page 200).

Estimating g from a hierarchical battery is therefore, mathematically, exactly the same problem as estimating any criterion, and can be done arithmetically in the same way. Because of the special nature of the hierarchical matrix of correlations, however, with its zero tetrad-differences, there is an easier way of calculating the estimate of g, due to Professor Spearman himself (*Abilities*, xviii). For its equivalence mathematically to the above see Appendix, paragraph 10.

Meanwhile we shall illustrate it by an example which will at least show that it is equivalent in this instance. The calculation is best carried out in tabular form, and is based entirely on the saturations or loadings of the tests with g, which are also their correlations with g.

Test	r_{ig}	r_{ig}^2	$1 - r_{ig}^2$	$\frac{r_{ig}^2}{1 - r_{ig}^2}$	$\frac{r_{ig}}{1 - r_{ig}^2}$	*Regression Coefficients* $\frac{1}{1+S} \times \frac{r_{ig}}{1 - r_{ig}^2}$
1	·9	·81	·19	4·2632	4·7368	·5533
2	·8	·64	·36	1·7778	2·2222	·2596
3	·7	·49	·51	·9608	1·3725	·1603
4	·6	·36	·64	·5625	·9375	·1095

$$S = 7{\cdot}5643$$
$$1 + S = 8{\cdot}5643$$
$$\frac{1}{1+S} = {\cdot}1168$$

The result, with much less calculation, is the same. The quantity S is of some importance in this formula. It is formed in the fourth column of the table, from which it will be seen that—

$$S = \frac{r_{1g}^2}{1 - r_{1g}^2} + \frac{r_{2g}^2}{1 - r_{2g}^2} + \ldots = \Sigma \frac{r_{ig}^2}{1 - r_{ig}^2}$$

It is clear that S will become larger and larger as the number of tests is increased.

Now, we saw that the square of the multiple correlation r_m is obtained when we multiply each of the weights by r_{ig} and sum the products. That is to say—

$$\begin{aligned} r_m^2 &= \Sigma\,(\text{weight} \times \text{saturation}) \\ &= \Sigma \left(\frac{1}{1+S} \cdot \frac{r_{ig}}{1 - r_{ig}^2} \times r_{ig} \right) \\ &= \frac{1}{1+S}\,\Sigma \frac{r_{ig}^2}{1 - r_{ig}^2} = \frac{S}{1+S} \end{aligned}$$

This fraction will be the nearer to unity, the larger S is ; and we can make S larger and larger by adding more and more (hierarchical) tests to the team. Thus in theory we can make a team to give as high a multiple correlation with g as we desire. It will also be noticed, however, from our table that the tests with high g saturation make much the largest contribution to S, and therefore to the multiple correlation.

2. *Estimating two factors simultaneously.*—We have seen in the preceding section how to estimate a man's g from his scores in a hierarchical team of tests, and in this we shall consider the broader question of estimating factors in general. Thus in Chapter V the four tests with correlations :

	1	2	3	4
1	.	·4	·4	·2
2	·4	.	·7	·3
3	·4	·7	.	·3
4	·2	·3	·3	.

were analysed into two common factors and four specifics with the loadings (see Chapter V, page 79).

	Common Factors *I*	*II*	*Specific Factors*			
1	·5164	.	·8563	.	.	.
2	·7746	·3162	.	·5477	.	.
3	·7746	·3162	.	.	·5477	.
4	·3873	.	.	.	.	·9220

Any one column of these loadings can be used as the criterion row in the calculation by Aitken's method, and the regression coefficients calculated with which to weight a man's test scores in order to estimate that factor for him. If, as is probable, we want to estimate both common factors, we can do the two calculations together, as shown on the next page. *Both* arrays of loadings are written below the matrix of intercorrelations, and then pivotal condensation automatically gives both sets of regression coefficients, with only one extra row in each slab of the calculation :

(1·0)	·4	·4	·2	−1·0	.	.	.	1·0
·4	1·0	·7	·3	.	−1·0	.	.	1·4
·4	·7	1·0	·3	.	.	−1·0	.	1·4
·2	·3	·3	1·0	.	.	.	−1·0	·8
·5164	·7746	·7746	·3873	.	.	.	.	2·5429
.	·3162	·3162	.	.	.	.	.	·6324
	(·84)	·54	·22	·40	−1·0	.	.	1·0
	1·00	·6429	·2619	·4762	−1·1905	.	.	1·1905
	·54	·84	·22	·40	.	−1·0	.	1·0
	·22	·22	·96	·20	.	.	−1·0	·6
	·5680	·5680	·2840	·5164	.	.	.	1·9365
	·3162	·3162	.	.		.	.	·6324
		(·4928)	·0786	·1429	·6429	−1·0	.	·3571
		1·0000	·1595	·2900	1·3046	−2·0292	.	·7246
		·0786	·9024	·0952	·2619	.	−1·0000	·3381
		·2028	·1352	·2459	·6762	.	.	1·2603
		·1129	−·0828	−·1506	·3764	.	.	·2560
			(·8899)	·0724	·1594	·1594	−1·0000	·2811
			1·0000	·0814	·1791	·1791	−1·1237	·3159
			·1029	·1871	·4116	·4116	.	1·1134
			−·1008	−·1833	·2291	·2291	.	·1742
				·1787	·3932	·3932	·1156	1·0809
				−·1751	·2472	·2472	−·1133	·2060

Regression Coefficients

If, therefore, we have a man's scores (in standard measure) in these four tests, our estimate of his Factor I will be—

$$\cdot1787 z_1 + \cdot3932 z_2 + \cdot3932 z_3 + \cdot1156 z_4$$

and estimates made in this way will have a multiple correlation r_m with the " true " values of the factor, in a number of different candidates, given by—

$$r_m^2 = \cdot1787 \times \cdot5164 + \cdot3932 \times \cdot7746 + \cdot3932 \times \cdot7746 + \cdot1156 \times \cdot3873 = \cdot7462$$

$$\therefore r_m = \cdot864$$

Similarly, the multiple correlation of the estimate of the second factor with the " true " values can be found to be—

$$r_m = \cdot 395$$

The two factors are not, therefore, estimated with equal accuracy by the team. As before, the whole calculation can be checked by a pooling square.

We have now found the regression equations for estimating the two common factors by treating each in turn as a " criterion." It is also possible to estimate a man's specific factors in the same way. Indeed, we might have written the loadings of the specific factors as four more rows below the common-factor loadings in the first slab and calculated their regression coefficients all in the one calculation. But it is easier to obtain the estimate of a man's specific by subtraction (compare *Abilities*, 1932 edition, page xviii, line 10). For example, we know that the second test score is made up as follows—

$$z_2 = \cdot 7746 f_1 + \cdot 3162 f_2 + \cdot 5477 s_2$$

where f_1 and f_2 are the man's common factors and s_2 his specific. We have estimated his f_1 and f_2, and we know his z_2; so we can estimate his s_2 from this equation. The estimates of *all* a man's factors, to be consistent with the experimental data, must satisfy this equation and similar equations for the other tests. If the estimate of the specific is actually made by a regression equation, just like the other factors, it will be found to satisfy this requirement.* From the estimates of *all* a man's factors, therefore, including any specifics, we can reconstruct his scores in the tests exactly. From only a few factors, however, even from all the common factors, we cannot reproduce the scores exactly, but only approximately.

3. *An arithmetical short cut* (Ledermann, 1938*a*, 1939*b*).—If the number of tests is *appreciably* greater than the number of common factors, the following scheme for

* It is interesting to note that we know the best *relative* loadings of the tests to estimate a specific by regression without needing to know how many common factors there are, or whether indeed any specific exists or not. (Wilson, 1934. For the same fact in more familiar notation, see Thomson, 1936*a*, 43.)

computing the regression coefficients will involve less arithmetical labour than the general formulæ expounded in Chapter XIV and applied to the factor problem in this chapter.*

For illustration, we shall use the data of the preceding section (page 225), although in that example the number of tests (four) exceeds the number of common factors (two) only by two, which is too small an amount to demonstrate fully the advantages of the present method. The common-factor loadings and the specifics of the four tests form a 4 × 2 matrix and a 4 × 4 matrix respectively, thus :

$$M_0 = \begin{bmatrix} \cdot 5164 & . \\ \cdot 7746 & \cdot 3162 \\ \cdot 7746 & \cdot 3162 \\ \cdot 3873 & . \end{bmatrix} \; ; \; M_1 = \begin{bmatrix} \cdot 8563 & & & \\ & \cdot 5477 & & \\ & & \cdot 5477 & \\ & & & \cdot 9220 \end{bmatrix}$$

the matrix M_0 being identical with the first two columns, and the matrix M_1 with the last four columns of the table on page 225. Before the data are subjected to the computational routine process, which will again consist in the pivotal condensation of a certain array of numbers, some preliminary steps have to be taken : (i) the loadings of each test are divided by the square of its specific, and the modified values are then listed in a new 4 × 2 matrix :

$$M_{01} = \begin{bmatrix} \cdot 7042 & . \\ 2\cdot 5820 & 1\cdot 0540 \\ 2\cdot 5820 & 1\cdot 0540 \\ \cdot 4556 & . \end{bmatrix}$$

e.g.

$$2\cdot 5820 = (\cdot 7746) \div (\cdot 5477)^2$$
$$1\cdot 0540 = (\cdot 3162) \div (\cdot 5477)^2$$

(ii) Next, the inner products (see footnote on page 74) of every column of M_0 in turn with every column of M_{01} are calculated and arranged in a 2 × 2 matrix :

* This short cut, in the form here given, is only applicable to orthogonal factors. For oblique factors, which are described in Chapter XII, modifications are necessary in Ledermann's formulæ, for which see Thomson (1949) and the later part of Section 19 of the Mathematical Appendix, page 365.

$$M'_0 M_{01} = J = \begin{bmatrix} 4{\cdot}5401 & 1{\cdot}6329 \\ 1{\cdot}6329 & {\cdot}6665 \end{bmatrix}$$

If there had been r common factors the matrix J would have been an $r \times r$ matrix. The arithmetic is simplified by the fact that J is always symmetrical about its diagonal, so that only the entries on and above (below) the diagonal need be calculated. (iii) Finally, each element on the diagonal of J is augmented by unity, giving, in the notation of matrix calculus, the matrix :

$$I + J = \begin{bmatrix} 5{\cdot}5401 & 1{\cdot}6329 \\ 1{\cdot}6329 & 1{\cdot}6665 \end{bmatrix}$$

This matrix is now " bordered " below by the matrix M_{01}, and on the right-hand side by a block of *minus ones* and *zeros* in the usual way. The process of pivotal condensation then yields the same regression coefficients as were obtained on page 226.

5·5401	1·6329	−1·0000	·	6·1730
1·0000	·2947	−·1805	·	1·1142
1·6329	1·6665	·	−1·0000	2·2994
·7042	·			·7042
2·5820	1·0540			3·6360
2·5820	1·0540			3·6360
·4556	·			·4556
	1·1853	·2947	−1·0000	·4800
	1·0000	·2486	−·8437	·4050
	−·2075	·1271		−·0804
	·2931	·4661		·7591
	·2931	·4661		·7591
	−·1343	·0822		−·0520
Regression Coefficients		·1787	−·1751	·0036
		·3932	·2473	·6404
		·3932	·2473	·6404
		·1156	−·1133	·0023

4. *Reproducing the original scores.*—Let us imagine a man who in each of the four tests in our example obtains a score of + 1 ; that is, one standard deviation above the average. We choose this set of scores merely to make the

With this table of loadings in our possession we might have given vocational advice to a man in a roundabout way. (Instead of inserting his scores in z_1, z_2, z_3, and z_4 in the equation for $\hat{z}_0$, we might have estimated his factors g, v, and F from his scores in the four tests, and then inserted these estimated factors in the specification equation of the occupation—

$$z_0 = \cdot 55g + \cdot 45v + \cdot 60F + \cdot 37s_0$$

(ignoring the specific s_0, which cannot be estimated from z_1, z_2, z_3, and z_4). *Had we done so, we should have arrived at exactly the same numerical estimate of his z_0 as by the direct method* (Thomson, 1936*a*, 49 and 50).)

The actual estimation of the factors g, v, and F from the four tests will form a good arithmetical exercise for the student. The beginning and end of the calculation of the regression coefficients is shown here, following exactly the lines of the smaller example on page 226 of this chapter :

								Check
1·00	·39	·69	·49	− 1	.	.	.	1·57
·39	1·00	·19	·27	.	− 1	.	.	1·26
·69	·19	1·00	·38	.	.	− 1	.	1·14
·49	·27	·38	1·00	.	.	.	− 1	·85
·66	·37	·52	·74	.	.	.	.	2·29
·52	.	·66	.	.	.	.	.	1·18
·21	·71	.	.	.	.	.	.	·92

This reduces by pivotal condensation step by step to the three sets of regression coefficients :

for $\hat{g}$	·300	·095	·095	·532
for $\hat{v}$	·353	− ·153	·581	− ·352
for $\hat{F}$	·121	·747	− ·148	− ·206

The result is to give us three equations for estimating g, v, and F from a man's scores in the four tests, viz.—

$$\hat{g} = \cdot 300z_1 + \cdot 095z_2 + \cdot 095z_3 + \cdot 532z_4$$
$$\hat{v} = \cdot 353z_1 - \cdot 153z_2 + \cdot 581z_3 - \cdot 352z_4$$
$$\hat{F} = \cdot 121z_1 + \cdot 747z_2 - \cdot 148z_3 - \cdot 206z_4$$

Now let us assume a set of scores z_1, z_2, z_3, z_4 for a man, and see what the estimate of his occupational ability is by the two methods, the one direct without using factors, the other by way of factors. Suppose his four scores are—

z_1	z_2	z_3	z_4
·2	·6	— ·4	·7

The estimates of his factors g, v, and F will therefore be—

$$\hat{g} = \cdot 300 \times \cdot 2 + \cdot 095 \times \cdot 6 + \cdot 095 \times (- \cdot 4) + \cdot 532 \times \cdot 7 = \quad \cdot 451$$
$$\hat{v} = \cdot 353 \times \cdot 2 - \cdot 153 \times \cdot 6 + \cdot 581 \times (- \cdot 4) - \cdot 352 \times \cdot 7 = - \cdot 500$$
$$\hat{F} = \cdot 121 \times \cdot 2 + \cdot 747 \times \cdot 6 - \cdot 148 \times (- \cdot 4) - \cdot 206 \times \cdot 7 = \quad \cdot 387$$

If now we insert these estimates of his factors into the specification equation of the occupation, ignoring its specific, we get for our estimate of his occupational success :

$$\hat{z}_0 = \cdot 55 \times \cdot 451 + \cdot 45 \times (- \cdot 500) + \cdot 60 \times \cdot 387 = \cdot 255$$

that is, we estimate that he will be about a quarter of a standard deviation better than the average workman. This by the indirect method using factors.

By the direct method, without using factors at all, we simply insert his test scores into the equation—

$$\hat{z}_0 = \cdot 390 z_1 + \cdot 431 z_2 + \cdot 222 z_3 + \cdot 018 z_4$$

and obtain—

$$\hat{z}_0 = \cdot 390 \times \cdot 2 + \cdot 431 \times \cdot 6 + \cdot 222 \times (- \cdot 4) + \cdot 018 \times \cdot 7$$
$$= \cdot 260$$

exactly the same estimate as before—for the difference in the third decimal place is entirely due to " rounding off " during the calculations. The third decimal place of the direct calculation is more likely to be correct, since it is so much shorter.

6. *Why, then, use factors at all?*—The reader may now ask, " What, then, is the use of estimating a man's factors at all ? " Well, in a case analogous to that of the present example it is quite unnecessary to use factors at all, and there is no doubt that a great many experimenters have rushed to factorial analysis with quite unjustifiable hopes of somehow getting more out of it than ordinary methods of vocational and educational advice can give without mentioning factors. But we must not go to the other extreme and " throw out the baby with the bath-water." There may be other reasons for using factors, apart from vocational advice. And even in giving such advice, which

tests and occupations into factors, still more the calculation of quantitative estimates of these factors, are as yet very inaccurate, and perhaps are inherently subject to uncertainty. A fluctuating and doubtful coinage can be a positive hindrance to trade, and barter may be preferable in such circumstances.

We showed in Section 5 above that a direct regression estimate of a man's ability in an occupation gives identically the same result as an estimate via the roundabout path of factors, so that at least when the direct regression estimate is possible there can be no quantitative advantage in using factors. When, however, is the direct regression estimate possible, and when is it impossible ?

To make the direct regression estimate we require the complete table of correlations of the tests with one another *and with the occupation*, and we have to know the candidate's scores in the tests. This implies that these same tests have been given to a number of workers whose proficiency in the occupation is known, for otherwise we would not know the correlations of the tests with the occupation. Under these ideal circumstances any talk of factors is certainly unnecessary so far as obtaining a quantitative estimate is concerned.

But suppose these ideal conditions do not hold ! These tests which we have given to the candidate have never been given, at any rate as a battery, to workers in the occupation, and their correlations with the occupation are unknown ! This situation is particularly likely to arise in vocational advice or guidance as distinguished from vocational selection. In the latter we are, usually on behalf of the employer, selecting men for a particular job, and we are practically certain to have tried our tests on people already in the job, and to be in a position to make a direct estimation without factors. But in vocational guidance we wish to gauge the young person's ability in very many occupations, and it is unlikely that just this battery of tests that we are using has been given to workers in all these different jobs. In that case we cannot make a direct regression estimate of our candidate's probable proficiency in every occupation. Can we, then, obtain an estimate in any other way ?

Other ways are conceivable, but it must at the outset be emphasized that *they are bound to be less accurate than the direct estimate without factors.* Although this battery of tests has not been given to workers in the occupation, perhaps other tests have, and by the aid of that other battery a factor analysis of the occupation has perhaps been made. If our tests enable the same factors to be estimated, we can gauge the man's factors and thence indirectly his occupational proficiency. Unfortunately, the "if" is a rather big one. Are factors obtained by the analysis of different batteries of tests the same factors; may they not be different even though given the same name? We shall discuss this very important point later, but meanwhile let us suppose that we have reasonable confidence in the identity of factors called by the same name by different workers with different batteries. Then the probable course of events would be something like this. An experimenter, using whatever tests he thinks practicable and suitable, analyses an occupation into factors. Another experimenter, at a different time and place, is asked to give advice to a candidate for that occupation. Using whatever tests he in his turn has available, he assesses in this candidate the factors which the previous experimenter's work leads him to think are necessary in the occupation, and gives his advice accordingly. The factors have played their part as a go-between, like a coinage. All depends on the confidence we have in the identity of the factors. We shall see later that there is only too much reason to think that the possibility of this confidence being misplaced has hardly been sufficiently realized by many over-enthusiastic factorists. And even if the common factors are identical, there remains the danger that the "specific" of the occupation may be correlated with some of the "specifics" of the tests, a fact which cannot be known unless the same tests have been given to workers in the occupation.

7. *Calculation of correlation between estimates.*—We said above that even although we make our analysis of the tests we use into uncorrelated factors, the *estimates* of these factors will be correlated, if we use communalities and thus have more factors than tests. Arithmetically, these

correlations are easily calculated from the inner products of (*b*), the loadings of the estimated factors with the tests (page 232), with (*a*), the loadings of the tests with the factors (page 231).

The matrix of loadings of the four tests with the three common factors is (page 231):

$$M = \begin{bmatrix} \cdot66 & \cdot52 & \cdot21 \\ \cdot37 & . & \cdot71 \\ \cdot52 & \cdot66 & . \\ \cdot74 & . & . \end{bmatrix}$$

and the matrix of the loadings of the three estimated factors with the four tests is (page 232):

$$N = \begin{bmatrix} \cdot300 & \cdot095 & \cdot095 & \cdot532 \\ \cdot353 & -\cdot153 & \cdot581 & -\cdot352 \\ \cdot121 & \cdot747 & -\cdot148 & -\cdot206 \end{bmatrix}$$

Then the matrix of variances and covariances of the estimated factors is—

$$K = NM$$

Performing the matrix multiplications as explained in Chapter X, Section 4, page 145, we obtain:

$$NM = \begin{bmatrix} \cdot300 & \cdot095 & \cdot095 & \cdot532 \\ \cdot353 & -\cdot153 & \cdot581 & -\cdot352 \\ \cdot121 & \cdot747 & -\cdot148 & -\cdot206 \end{bmatrix} \begin{bmatrix} \cdot66 & \cdot52 & \cdot21 \\ \cdot37 & . & \cdot71 \\ \cdot52 & \cdot66 & . \\ \cdot74 & . & . \end{bmatrix}$$

$$= \begin{bmatrix} \cdot676 & \cdot219 & \cdot130 \\ \cdot218 & \cdot567 & -\cdot034 \\ \cdot127 & -\cdot035 & \cdot556 \end{bmatrix} = K$$

If our arithmetic throughout the whole calculation of these loadings had been perfectly accurate, the matrix K would have been perfectly symmetrical about its diagonal. The actual discrepancies (as ·127 and ·130) are a measure of the degree of arithmetical accuracy attained.

The matrix K thus arrived at gives by its diagonal elements ·676, ·567, and ·556, the *variances* of the three estimated factors (that is, the squares of their standard deviations), and by its other elements their *covariances* in pairs (that is, their overlap with one another). The correlation of any two estimated factors is equal to (see Chapter I, Figure 2)—

$$r_{ij} = \frac{\text{covariance } (ij)}{\sqrt{\text{variance } (i) \times \text{variance } (j)}}$$

From K we can therefore form the matrix of correlation of the estimated factors. It is :

$$\begin{bmatrix} 1{\cdot}000 & {\cdot}353 & {\cdot}212 \\ {\cdot}353 & 1{\cdot}000 & -{\cdot}061 \\ {\cdot}212 & -{\cdot}061 & 1{\cdot}000 \end{bmatrix}$$

wherein ·353, for example, is $\cdot 219 \div \sqrt{(\cdot 676 \times \cdot 567)}$. Although, therefore, the " true " factors g and v are uncorrelated, their estimates $\hat{g}$ and $\hat{v}$ are correlated to an amount ·353. The " true " factors g, v, and F are in standard measure, but their estimates $\hat{g}$, $\hat{v}$, and $\hat{F}$ have variances of only ·676, ·567, and ·556 instead of unity. These variances, be it noted in passing, are equal also to the squares of the correlations between g and $\hat{g}$, v and $\hat{v}$, F and $\hat{F}$.

Not only are the estimates of the common factors correlated among themselves ; they are correlated with the specifics, so that the *estimates* of the specifics are not strictly specific. As a numerical illustration we may take the hierarchical matrix used in Section 1, pages 221 *ff*.

	z_1	z_2	z_3	z_4
z_1	1·00	·72	·63	·54
z_2	·72	1·00	·56	·48
z_3	·63	·56	1·00	·42
z_4	·54	·48	·42	1·00

The regression estimate of g from this battery is, as we found on page 223)—

$$\hat{g} = \cdot 553 z_1 + \cdot 259 z_2 + \cdot 160 z_3 + \cdot 109 z_4$$

The regression estimates for the four specifics can also be found, either by a full calculation like that of page

226, or by the simpler method of subtraction of page 227. Thus, to estimate s_1 in our present example we know that—

$$z_1 = \cdot 9g + \sqrt{1 - \cdot 9^2}\, s_1$$
$$= \cdot 9g + \cdot 436 s_1$$

Also we know that the estimates $\hat{g}$ and $\hat{s}_1$ will satisfy the same equation—

$$z_1 = \cdot 9\hat{g} + \cdot 436\hat{s}_1$$

that is—

$$\hat{s}_1 = \frac{z_1 - \cdot 9\hat{g}}{\cdot 436}$$

On inserting the expression for $\hat{g}$ into this we get—

$$\hat{s}_1 = 1\cdot 152 z_1 - \cdot 535 z_2 - \cdot 333 z_3 - \cdot 225 z_4$$

and similarly—

$$\hat{s}_2 = -\cdot 737 z_1 + 1\cdot 313 z_2 - \cdot 215 z_3 - \cdot 145 z_4$$
$$\hat{s}_3 = -\cdot 542 z_1 - \cdot 253 z_2 + 1\cdot 242 z_3 - \cdot 106 z_4$$
$$\hat{s}_4 = -\cdot 415 z_1 - \cdot 194 z_2 - \cdot 121 z_3 + 1\cdot 169 z_4$$

We have now both N, the matrix of loadings of the *estimated* factors $\hat{g}$, $\hat{s}_1$, $\hat{s}_2$, $\hat{s}_3$, $\hat{s}_4$ with the four tests, and M, which we already know, the matrix of loadings of the four tests with the five factors g, s_1, s_2, s_3, and s_4, namely :

$$M = \begin{bmatrix} \cdot 9 & \cdot 436 & . & . & . \\ \cdot 8 & . & \cdot 600 & . & . \\ \cdot 7 & . & . & \cdot 714 & . \\ \cdot 6 & . & . & . & \cdot 800 \end{bmatrix}$$

From their product NM we obtain the matrix K of variances and covariances of the estimated factors, namely :

$$\begin{bmatrix} \cdot 553 & \cdot 259 & \cdot 161 & \cdot 109 \\ 1\cdot 152 & -\cdot 535 & -\cdot 333 & -\cdot 225 \\ -\cdot 737 & 1\cdot 313 & -\cdot 215 & -\cdot 145 \\ -\cdot 542 & -\cdot 253 & 1\cdot 242 & -\cdot 106 \\ -\cdot 415 & -\cdot 194 & -\cdot 121 & 1\cdot 169 \end{bmatrix} \begin{bmatrix} \cdot 9 & \cdot 436 & . & . & . \\ \cdot 8 & . & \cdot 600 & . & . \\ \cdot 7 & . & . & \cdot 714 & . \\ \cdot 6 & . & . & . & \cdot 800 \end{bmatrix}$$

$$= \begin{bmatrix} \cdot 880 & \cdot 241 & \cdot 155 & \cdot 115 & \cdot 087 \\ \cdot 241 & \cdot 502 & -\cdot 321 & -\cdot 238 & -\cdot 180 \\ \cdot 150 & -\cdot 321 & \cdot 788 & -\cdot 154 & -\cdot 116 \\ \cdot 116 & -\cdot 236 & -\cdot 152 & \cdot 887 & -\cdot 085 \\ \cdot 088 & -\cdot 181 & -\cdot 116 & -\cdot 086 & \cdot 935 \end{bmatrix} = K$$

Again, we have a check on the accuracy of our arithmetic, for K will, if we have been accurate, be exactly symmetrical about its principal diagonal, i.e. its diagonal running from north-west to south-east. The largest discrepancy in our case is between ·150 and ·155. Moreover, since in this case K includes *all* the factors, we have another check which was not available when we calculated a K for common factors only: the sum of the elements in the principal diagonal (called the "trace," or in German the "*Spur*") here must come out equal to the number of tests. In our case we have—

$$\cdot 880 + \cdot 502 + \cdot 788 + \cdot 887 + \cdot 935 = 3 \cdot 992$$

and there are four tests. These elements which form the trace of K are, it will be remembered, the variances of the estimates $\hat{g}$, $\hat{s}_1$, $\hat{s}_2$, $\hat{s}_3$, and $\hat{s}_4$. So that we see that the total variances of the *five* factors is no greater than the total variance (viz. 4) of the *four* tests in standard measure. This is only another instance of the general law that we cannot get more out of anything than we put into it (at any rate, not in the long run).

From K we can at once calculate the correlation of the estimated factors. Adjusting the slight arithmetical departures from symmetry, we get:

	$\hat{g}$	$\hat{s}_1$	$\hat{s}_2$	$\hat{s}_3$	$\hat{s}_4$
$\hat{g}$	1·000	·362	·184	·131	·096
$\hat{s}_1$	·362	1·000	− ·510	− ·354	− ·263
$\hat{s}_2$	·184	− ·510	1·000	− ·183	− ·135
$\hat{s}_3$	·131	− ·354	− ·183	1·000	− ·094
$\hat{s}_4$	·096	− ·263	− ·135	− ·094	1·000

from which we see that g is correlated with each of the estimated specifics positively, while the latter are correlated negatively among themselves, in this (a hierarchical) example.

We have then this result, that although we set out to analyse our battery of tests into independent uncorrelated factors, the estimates which we make of these factors are correlated with one another, and instead of being in standard measure have variances, and therefore standard deviations, less than unity. We could, of course, make them unity by dividing all our estimates by their calculated standard deviation. But that would make no change in their correlations.

The cause of all this is the excess of factors over tests, and consequently this drawback—the correlation of the estimates—depends upon the ratio of the number of factors to the number of tests. The extra factors are the common factors, for there is a specific to each test, and therefore with the same number of common factors the correlation between the estimates will decrease as the number of tests in the battery increases. Just as in the hierarchical case one of the tasks of the experimenter is to find tests to add to the number in his battery without destroying its hierarchical nature, so in the case of a battery which can be reduced to rank 2, 3, 4 . . . or r, a task will be to add tests to the battery which with suitable communalities will leave the rank unchanged and the pre-existing communalities unaltered, in order that the common factors may be the more accurately estimated, and the estimates be more nearly uncorrelated.

8. *Bartlett's method of estimation.*—M. S. Bartlett (1935, 1937*a*, 1938) has proposed to estimate the common factors, not by the ordinary regression method used above, but by a method which minimizes the sum of the squares of a man's specific factors (already, however, maximized by the principle of using as few common factors as possible).

The way in which Bartlett's estimates differ from regression estimates of factors can be very clearly seen by thinking in terms of the geometrical picture already used in earlier chapters. When the factors outnumber the tests, the vectors representing the former are in a space of higher dimensions than the test space.

The individual person is represented in the test space by a point, namely that point P whose projections on to the test vectors give his test scores. We do not know a representative *point* for this individual in the complete factor space, however. His representative point Q may be, for all we know, *anywhere* in the subspace which is perpendicular to the test space and intersects with it at P. In these circumstances the regression method takes refuge in the assumption that this individual is average in all qualities of which we know *nothing*; that is, in all qualities orthogonal to our test space. It therefore assumes P to be his point also in the factor space, and projects P on to the factor axes to get the estimates of his factors.

Bartlett's method is equivalent to a different assumption about the position of the point Q. Within the complete factor space there is a subspace which contains the *common* factors. Of all the positions open to the point Q, Bartlett's method chooses that one which is nearest to the common-factor space, and from thence projects on to the common-factor vectors. This is equivalent to making the assumption that this man is *not* average in the qualities about which we know nothing, but instead possesses in those unknown qualities just those degrees of excellence which bring his representative point to the chosen point Q. Because men are most frequently near the average, the regression assumption is more likely.

9. *The geometrical interpretation of Bartlett's method.*—All this can be most clearly seen (because a perspective diagram can be made) in the case of estimating one genera factor g only, the hierarchical case. A figure like Figure 30 will illustrate this case, if we take y and z there to be two tests and x to be the g vector (see page 214).

The man's representative point in the yz plane is P. But we do not know his representative point Q in solid three-dimensional space, only that it is somewhere on the line $P'PP''$. The regression method assumes that it is actually at P, the average, and projects P itself on to the g line to get the estimate OL of g. Bartlett's method, on the other hand, assumes that Q is at that point on $P'PP''$ where it most nearly approaches the g line, that is, somewhere near the position Q in our diagram. Bartlett's estimate of g is then represented by OL'.

Now, any point on the line $P'PP''$, when projected on to the test vectors y and z, gives the same two test scores. There is, in general, no point on the line g which does this exactly. But clearly L', of all the points on g, will be the point whose projections most nearly fall on Y and Z, for X' is as near as possible to the line $P'PP''$. That is, the projection of X' on to the plane of the tests falls as near to the point P as is possible. In other words, if we ignore the specifics entirely and use only the estimated g in the specification of y and z, Bartlett's estimate comes as near as is possible to giving us back the full scores OM and ON. If the regression estimate OL is projected on to the lines y and z, it will obviously give a worse approximation.

The regression method, in order to recover as much as possible of the original scores, would have to make a second estimate of them. For the estimates of g represented by quantities like OL are not in standard measure. Before projecting the point L on to the lines y and z, therefore, to recover the original scores as far as possible, the regression method would alter the scale of its space along the g vector until the quantities like OL *were* in standard measure. This would not only change the position of L on the line, it would change the angles which the lines in the figure make with one another ; *and would change them exactly in such a manner that, in the new space, the projection of* OL *on to* y *and* z *would fall exactly where the Bartlett projections from* L′ *fall in the present space* (Thomson, 1938*a*).

There is, therefore, no final difference in excellence between the two methods in the matter of restoring the original scores as fully as possible, but the regression method takes two bites at the cherry. On the other hand, the regression estimates can be put straight into the specification equation of an occupation which is known to require just these common factors, whereas here it is the Bartlett method which has to have a second shot.

Both methods have to change their estimate of g when a new test is added to the battery. For the man is not very likely to have, in the specific of this new test, either the average value previously assumed by the regression method, or the special value assumed by the Bartlett method. But he is more likely to have the former than the latter, so the Bartlett estimates will change more than do the regression estimates as the battery grows. Ultimately, when the number of tests becomes infinite, the two forms of estimate will agree.

In the case of estimates of one general factor g from a hierarchical battery, the Bartlett estimates differ from the regression estimates only in scale. They put the candidates in the same order of merit for g as do the regression estimates, but give them a greater scatter, making the high g's higher and the low g's lower. The formula is—

$$\frac{1}{S}\Sigma\frac{r_{ig}z_i}{1-r_{ig}^2}$$

instead of Spearman's—

$$\frac{1}{1+S}\Sigma\frac{r_{ig}z_i}{1-r_{ig}^2}\text{ (see page 224).}$$

With more than one common factor, the connexion between the two kinds of estimate is not so simple (Appendix, Section 13). The mathematical reader will be able to calculate the Bartlett factor estimates from the matrix formulæ given in the Appendix.

10. *Estimation of oblique factors.*—In applying the method of Section 2 to oblique factors, it is important to note that we must use, below the matrix of correlations of the tests, in a calculation like that on page 226, the matrix of correlations of the primary factors with the tests. These are the elements of the *structure on the primary factors*, $F(\Lambda')^{-1}D$, transposed so that columns become rows and *vice versa*. It would not do to use the structure on the reference vectors, which is all that most experimenters content themselves with calculating.

Ledermann's short cut (Section 3 above) requires considerable modification in the case of oblique factors. See Thomson (1949) and the later part of Section 19 of the Mathematical Appendix, page 365.

PART V

CORRELATIONS BETWEEN PERSONS

CHAPTER XVI

REVERSING THE RÔLES*

1. *Exchanging the rôles of persons and tests.*—In all the previous chapters the correlations considered have been correlations between tests, and the experiments envisaged were experiments in which comparatively few tests were administered to a large number of persons. For each test there would, therefore, be a long list of marks. The whole set of marks would make an oblong matrix, with a few *rows* for the tests, and a very large number of *columns* for the persons—we will choose that way of writing it, of the two possibilities.

From such a set of marks we then calculated the correlation coefficients for each pair of tests, and our analysis of the tests into factors was based upon these. In the process of calculating a correlation coefficient we do such things to the row of marks in each test as finding its average, and finding its standard deviation. We quite naturally assume that we can legitimately carry out these operations. We assume, that is, that in the row of marks for one test these marks are comparable magnitudes which at any rate rise and fall with some mental quality even if they do not strictly speaking measure it in units, like feet or ounces.

The question we are going to ask in this part of this book is whether, in the above procedure, the rôles of persons and of tests can be exchanged (Thomson, 1935*b*, 75, Equation 17), and if so what light this throws upon factorial analysis. Instead of comparatively few tests

* The first explicit reference to correlations between persons in connexion with factor technique seem to have been made independently and almost simultaneously by Thomson (1935*b*, July) and Stephenson (1935, August), the former being pessimistic, the latter optimistic. But such correlations had actually been used much earlier by Burt and by Thomson, and almost certainly by others. See Burt and Davies, *Journ. Exper. Pedag.*, 1912, **1**, 251.

(perhaps two or three dozen) and a very large number of persons, suppose we have comparatively few persons, and a large number of tests, and find the correlations between the persons. In that case our matrix of marks would be oblong in the other direction, with a large number of rows for the tests, and a small number of columns for the persons, and each correlation, instead of being as before between two rows, would be between two columns. Taking only small numbers for purposes of an explanatory table, we would have in the ordinary kind of correlations a table of marks like this :

				Persons			
	×	×	×	×	×	×	×
Tests	×	×	×	×	×	×	×
	×	×	×	×	×	×	×

while for correlations between persons we would have a table of marks like this :

		Persons	
	×	×	×
	×	×	×
	×	×	×
Tests	×	×	×
	×	×	×
	×	×	×
	×	×	×

But we meet at once with a serious difficulty as soon as we attempt to calculate a correlation coefficient between two persons from the second kind of matrix. To do so, we must find the average of each column, just as previously we found the average of each row for the other kind of correlation. But to find the average of each column (by adding all the marks in that column together and dividing by their number) is to assume that these marks are in some sense commensurable up and down the column, although each entry is a mark for a different test, on a scoring system which is wholly arbitrary in each test (Thomson, 1935*b*, 75–6).

To make this difficulty more obvious, let us suppose that the first four tests are :

1. A form-board test ;
2. A dotting test ;
3. An absurdities test ;
4. An analogies test.

In each of these the experimenter has devised some kind of scoring system. Perhaps in the form-board test he gives a maximum of 20 points, and in the dotting test the score may be the number of dots made in half a minute. But to find the average of such different things as this is palpably absurd, and the whole operation can be entirely altered by an arbitrary change like taking the number of seconds to solve the form board instead of giving points.

2. *Ranking pictures, essays, or moods.*—This is a very fundamental difficulty which will probably make correlations between persons in the general case impossible to calculate. In certain situations, however, it does not arise, namely where each person can put the " tests " in an order of preference according to some criterion or judgment (Stephenson, 1935), and it is with cases of this kind that we shall deal in the first place. Usually the " tests " here are not really different tests like those named above, but are perhaps a number of children's essays which have to be placed in order of merit, or a number of pictures in order of æsthetic preference, or a number of moods which the subject has to number, indicating the frequency of their occurrence in himself. Indeed, the subject might not only give an order of preference to, say, the essays, but might give them actual marks, and there would be no absurdity in averaging the column of such marks, or in correlating two such columns, made by different persons.

Such a correlation coefficient would show the degree of resemblance between the two lists of marks given to the children, or given to a set of pictures according to their æsthetic value. It would indicate, therefore, a resemblance between the minds of the two persons who marked the essays or judged the pictures. A matrix of correlations between several such persons might look exactly like the matrices of correlations between tests, and could be analysed in any of the same ways. What would the "factors" which resulted from such an analysis mean when the correlations were between persons? Take an imaginary hierarchical case first.

3. *The two sets of equations.*—In test analysis the common factor found was taken to be something called into play by each test, the different tests being differently loaded with it. The test was represented by an equation such as—

$$z_4 = \cdot 6g + \cdot 8s_4$$

For each of the numerous persons who formed the subjects of the testing, an estimate was made of *his* g, and another estimate could be made of *his* s_4. The different tests were combined into a weighted battery for this purpose of estimating a man's amount of g. His score in Test 4 would then be made up of his g and s_4 inserted in the above specification equation.

$$z_{4\cdot 9} = \cdot 6g_9 + \cdot 8s_{4\cdot 9}$$

would be the score of the ninth person in Test 4.

By analogy, when we analyse a matrix consisting of correlations between persons, we arrive at a set of equations describing the persons in terms of common and specific factors. Corresponding to a hierarchical battery of tests, we could conceivably have a hierarchical team of persons, from which we would exclude any person too similar to one already included. Each person in the hierarchical team would then be made up of a factor he shared with everyone else in the team, and a specific factor which was his own idiosyncrasy. An equation like—

$$z_9 = \cdot 4g' + \cdot 917s_9'$$

would now specify the composition of the ninth person. g' is something all the persons have, s_9' is peculiar to Person 9. The *loadings* now describe the person, and the amount of g' "possessed" or demanded by each test can be estimated by exactly the same techniques employed in Chapter XV. The score which Test 4 would elicit from Person 9 would be obtained by inserting the g' and s_9'

" possessed " by that test into the specification equation of Person 9, giving—

$$z_{9\cdot4} = \cdot4g_4' + \cdot917s_{9\cdot4}'$$

This equation is to be compared with the former equation—

$$z_{4\cdot9} = \cdot6g_9 + \cdot8s_{4\cdot9}$$

Both equations ultimately describe the same score, but $z_{9\cdot4}$ is not identical with $z_{4\cdot9}$. The raw score X is the same, but the one standardized z is measured from a different zero, and in different units, from the other. Disregarding this for the moment, we see that with the exchange of rôles of tests and persons, *the loadings and the factors have also changed rôles.* Formerly, persons possessed different amounts of g, and tests were differently loaded with it. Now, tests possess different amounts of g', and persons are differently loaded with it. We feel impelled to inquire further into the relationships of these complementary factors and loadings.

The test which is most highly saturated with g is that one which, in terms of Spearman's imagery, requires most expenditure of general mental energy, and is least dependent upon specific neural engines. It correlates more with its fellow-members of the hierarchical battery than any other test among them does. It represents best what is common to them all.

The man, in a hierarchical team of men, who is most highly saturated with g' is that man who is most like all the others. His correlations with them are higher than is the case for any other man in the team. He is the individual who best represents the type. But a nearer approach to the type can be made by a weighted team of men, just as formerly we weighted a battery of tests to estimate their common factor.

4. *Weighting examiners like a Spearman battery.*—Correlations of this kind between persons were used long before any idea of what Stephenson has called " inverted factorial analysis " was present. The author and a colleague found in the winter of 1924–5 a number of correlations between experienced teachers who marked the essays written by fifty schoolboys upon " Ships " (Thomson and Bailes,

1926). One table or matrix of such correlations between the class teacher and six experienced head masters who marked the essays independently of one another, was as follows :

	Te	*A*	*B*	*C*	*D*	*E*	*F*
Te	.	·60	·69	·56	·69	·63	·67
A	·60	.	·53	·50	·54	·55	·68
B	·69	·53	.	·60	·65	·66	·64
C	·56	·50	·60	.	·67	·67	·65
D	·69	·54	·65	·67	.	·54	·69
E	·63	·55	·66	·67	·54	.	·69
F	·67	·68	·64	·65	·69	·69	.

In the article in question, these different markers were compared by correlating each with the pool of all the rest. These correlations are shown in the first row of the table below.

Purely as an illustrative example, let us make also an approximate analysis of this matrix, and take out at any rate its chief common factor. On the assumption that it is roughly hierarchical, we can use Spearman's formula—

$$\text{Saturation} = \sqrt{\left\{\frac{A^2 - A'}{T - 2A}\right\}}\,^*$$

More easily we can insert its largest correlation coefficient as an approximate communality for each test, and find Thurstone's approximate first-factor loadings (see Chapter V, page 70). We get for the saturations or loadings the second and third rows of this table :

	Te	*A*	*B*	*C*	*D*	*E*	*F*
Correlation with pool of rest	·77	·67	·76	·73	·76	·75	·82
Spearman saturations	·814	·704	·796	·766	·798	·788	·861
Thurstone method	·81	·73	·80	·78	·80	·80	·85

We see that *F* is the most " typical " examiner of these essays, in the sense that he is more highly saturated with what is common to all of them ; while *A* conforms least to the herd.

With the same formula which on page 224 we used to estimate a man's *g* from his test scores, we could here estimate

* See Chapter III, page 43.

an essay's g' from its examiner scores. That is to say, the marks given by the different examiners would be weighted in proportion to the quantities—

$$\frac{\text{Saturation with } g'}{1 - \text{saturation}^2}$$

where g' is that quality of an essay which makes a common appeal to all these examiners. Their marks (after being standardized) would therefore be weighted in the proportions $\cdot 814/(1 - \cdot 814^2)$, etc., that is :

	Te	*A*	*B*	*C*	*D*	*E*	*F*
	2·41	1·40	2·17	1·85	2·20	2·08	3·33
or	·72	·42	·65	·56	·66	·63	1·00

to make global marks for the essays, which could then be reduced to any convenient scale. If this were done, the result would be the " best " estimate* of that aspect or set of aspects of the essay which all these examiners are taking into account, disregarding all that can possibly be regarded as idiosyncrasies of individual examiners. Whether we think it the best estimate in other senses is a matter of subjective opinion. We may wish the " idiosyncrasies " (the specific, that is) of a certain examiner to be given great weight. It clearly would not do, for example, to exclude Examiner *A* from the above team *merely* because he is the most different from the common opinion of the team, without some further knowledge of the men and the purpose of the examination. The " different " member in a team might, for example, be the only artist on a committee judging pictures, or the only Democrat in a court judging legal issues, or the only woman on a jury trying an accused girl. But in non-controversial matters, if all are of about equal experience, it is probable that this system of weighting, restricting itself to what is certainly common to all, will be most generally acceptable as fairest.

* Best whether we adopt the regression principle or Bartlett's. For if only one " common factor " is estimated, the difference is one of unit only, and the weighting in the text is the " best " on both systems.

5. *Example from "The Marks of Examiners."*—This form of weighting examiners' marks has probably never yet been used in practice. But it has been employed, by Cyril Burt, in an inquiry into the marks given by examiners (Burt, 1936). As an example, we take the marks given independently by six examiners to the answer papers of fifteen candidates aged about 16, in an examination in Latin. (The example is somewhat unusual, inasmuch as these candidates were a specially selected lot who had all been adjudged equal by a previous examiner, but it will serve as an illustration if the reader will disregard that fact.) The marks were (*op. cit.*, 20) :

Cand.	*A*	*B*	*C*	*D*	*E*	*F* *Examiners*
1	39	43	52	37	43	40
2	39	44	50	43	43	46
3	44	51	55	47	46	46
4	37	46	43	44	40	43
5	38	47	55	35	43	45
6	45	50	54	45	45	49
7	42	52	51	45	44	46
8	43	49	53	47	46	46
9	32	42	49	34	36	38
10	37	40	48	37	39	42
11	38	42	47	39	36	39
12	40	44	50	41	36	42
13	38	43	50	36	34	41
14	35	45	49	37	40	40
15	32	38	41	28	34	34

The correlations between the examiners calculated from this table are (the examiner with the highest total correlation leading) :

	F	*A*	*B*	*E*	*D*	*C*
F	.	·86	·84	·82	·84	·71
A	·86	.	·80	·74	·85	·71
B	·84	·80	.	·80	·81	·67
E	·82	·74	·80	.	·72	·69
D	·84	·85	·81	·72	.	·48
C	·71	·71	·67	·69	·48	.

If, assuming this table to be hierarchical, we find each examiner's saturation with the common factor by Spearman's formula, we obtain (with Professor Burt, *op. cit.*, 294) :

F	*A*	*B*	*E*	*D*	*C*
·95	·92	·91	·87	·84	·72

In the sense, therefore, of being most typical, *F* is here the best examiner. The proportionate weights to be given to each examiner, in making up that global mark for the candidate which will best agree with the common factor of the team of examiners, are, as before—

$$\frac{\text{Saturation}}{1 - \text{saturation}^2}$$

provided the marks have first been standardized. The resulting weights, giving *F* the weight unity, are :

F	*A*	*B*	*E*	*D*	*C*
1·00	·61	·54	·37	·29	·15

(If the weights are to be applied to the raw or unstandardized marks, they must each be divided by that examiner's standard deviation.)

The marks thus obtained are only an estimate of the "true" common-factor mark for each child, just as was the case in estimating Spearman's *g* ; and the correlation of these estimates with the "true" (but otherwise undiscoverable) mark will be, as there (Chapter XV, page 224)—

$$r_m = \sqrt{\frac{S}{1 + S}}$$

where S is the sum of all the six quantities—

$$\frac{\text{Saturation}^2}{1 - \text{saturation}^2}$$

In our case this gives—

$$r_m = \cdot 98$$

The best examiner's marking itself correlated with the hypothetical "true" mark to the amount ·95, so that the improvement is not worth the trouble of weighting, especially as the simple average of the team of examiners gives ·97. But in some circumstances the additional labour might be worth while, and there is an interest in knowing which examiners conform least and which most to the team, and having a measure of this.

After the saturation of each examiner with the hypothetical common factor has been found, the correlations due to that factor can be removed from the table exactly as in analysing tests. The residues, as there, may show the presence of other factors ; and "specific" resemblances or antagonisms between pairs of examiners, or minor factors running through groups of examiners, may be detected and estimated.

In short, all the methods used on correlations between tests may be employed on correlations between examiners. The tests have come alive and are called examiners, that is all. But since the child's performance, judged by the different examiners differently, is here nevertheless the same identical performance, our interpretation of the results is different. The two cases throw light on one another. A Spearman hierarchical battery of tests may estimate each child's general intelligence, which is there something in common among the tests. The examiners may have been instructed to mark exclusively for what they think is general intelligence. In that case their weighted team will estimate for each child a general intelligence, which is something in common among the somewhat discrepant ideas the examiners hold on this matter.

6. *Preferences for school subjects.*—In the previous sections we have discussed correlations between examiners who all mark the same examination papers. The purpose of their marking these papers is to award prizes, distinctions, passes, and failures to the candidates. The examiners are a means to this end ; the reason for employing several of them is to obtain a list of successes and failures in which we can have greater confidence. The technique described is one which enables us to combine their marks, on certain assumptions, to greatest advantage. But it can, as in the inquiries described in *The Marks of Examiners*, be turned to compare individual examiners, and to evaluate the whole process of examining.

It is only a step to another, very similar, experiment in which objects evaluated by the " examiners " are not the works of candidates in an examination, but are objects chosen for the express purpose of gaining an insight into the minds of those asked to judge them. Thus we might ask several persons each to evaluate on some scale the æsthetic appeal of forty or fifty works of art (Stephenson, 1936*b*, 353), or ask a number of school pupils each to place in order of interest a list of school subjects.

Stephenson (1936*a*) asked forty boys and forty girls attending a higher school in Surrey, England, thus to place in order of their preference twelve school subjects represented by sixty examination papers, and calculated for about half these pupils the correlation coefficients between them. To explain the kind of outcome that may be expected from such an experiment it will be sufficient for us to quote his data for a smaller number of pupils, say eight girls, avoiding anomalous cases for simplicity in a first consideration. The correlations between them were as follows (*op. cit.*, 50) :

Girl	3	4	5	7	17	18	19	20
3	.	·59	·31	·26	—·02	—·16	—·38	—·35
4	·59	.	·75	·42	—·23	—·01	—·66	—·03
5	·31	·75	.	·65	—·29	—·02	—·18	—·08
7	·26	·42	·65	.	—·50	—·15	—·54	—·17
17	—·02	—·23	—·29	—·50	.	·60	·52	·72
18	—·16	—·01	—·02	—·15	·60	.	·09	·79
19	—·38	—·66	—·18	—·54	·52	·09	.	·40
20	—·35	—·03	—·08	—·17	·72	·79	·40	.

This table at once suggests that these girls fall into two types. Girls 3, 4, 5, and 7 correlate positively among themselves ; they have somewhat similar preferences among school subjects. Girls 17, 18, 19, and 20 correlate positively among *them*selves. But the two groups correlate negatively with one another. The two types were different in their order of preference, Type I tending, for example, to put English and French higher, and Physics and Chemistry lower, than Type II (though both were agreed that Latin was about the least lovable of their studies!).

7. *A parallel with a previous experiment.*—This experiment, it will be seen, forms a parallel to that inquiry (also by Stephenson) described in Chapter I, Section 9, where *tests* fell into two types, verbal and pictorial, with correlations falling there as here into four quadrants. If we call the two types of school pupil here the linguistic (L) and the scientific (S), and again use C for the cross-correlations, the diagram corresponding to that on page 16 of Chapter I is :

L	C
C	S

The chief difference between the two cases is that there the cross-correlations, though smaller than hierarchical order in the whole table would demand, were nevertheless positive. Here, however, the cross-correlations are actually negative.

It is true that the signs of all the correlations in the C quadrants can in either case be reversed, by reversing the order of the lists either of all the earlier or all the later variables (there tests, here pupils). But that is not really permissible in either case. We have no doubt which is the top and which the bottom end of a list of marks, whether in a verbal test or a pictorial test ; and to reverse the order of preference given by either the linguistic or the scientific pupils would be simply to stultify the inquiry. There is, therefore, a real difference between the cases. In the present set of correlations something is acting as an " interference factor."

In Chapter I we explained the correlations and their tetrad-differences by the hypothesis of three uncorrelated factors g, v, and p required in various proportions by the tests, and possessed in various amounts by the children. The *loadings* which indicated the proportions of the factors in each test we tacitly assumed to be all positive. Thurstone expressly says that it is contrary to psychological expectation to have more than occasional negative loadings.

8. *Negative loadings.*—Let us endeavour to make at least a qualitative scheme of factors to express the correlations between the pupils, factors possessed in various amounts by the subjects of the school curriculum, and demanded in various proportions by each pupil before he will call the subject interesting. One type of pupil weights heavily the linguistic factor in a subject in evaluating its interest to him. The other type weights heavily the scientific factor in a subject in judging its attraction for him. But to explain actual negative correlations between pupils we must assume that some of the loadings are negative, assume, that is, that some of the children are actively repelled by factors which attract others. Common sense does not think thus. Common sense says that two children may put the subjects in opposite orders, even though they both like them all, provided they don't like them equally well. But then common sense is not anxious to analyse the children into *uncorrelated additive* factors. If each child is thus expressed as the weighted sum of various factors, two children can correlate negatively only if some of the loadings are negative in the one child and positive in the other, for the correlation is the inner product of the loadings. Since Stephenson has found numerous negative correlations between persons, and since few negative correlations are reported between tests, we seem here to have an experimental difference between the two kinds of correlation, and if ever correlations between persons come to be analysed as minutely and painstakingly as correlations between tests, it would seem that the free admission of negative loadings would be necessary.* The present matrix can in fact be roughly analysed into two general factors, one of which has positive loadings in all pupils, while the other is positively loaded in the one type, negatively loaded in the other.

9. *An analysis of moods.*—A still more ingenious application by Stephenson of correlations between persons is in an experiment in which for each person a " population "

* See Stephenson, 1936*b*, 349.

of thirty *moods*, such as " irascible," " cheerful," " sunny," were rated for their prevalence and intensity for each of ten patients in a mental hospital, and for six normal persons (Stephenson, 1936*c*, 363). This time the correlation table indicated three types, corresponding to the manic-depressives, the schizophrenes, and the normal persons, each type correlating positively within itself, but negatively or very little with the other types. These experiments were only illustrative, and it remains to be seen whether factors which will prove acceptable psychologically will be isolated in persons in the same manner as *g*, and the verbal factor, have been isolated in tests. The parallel between the two kinds of correlation and analysis is, however, certainly likely to throw light on the nature of factors of both kinds.

CHAPTER XVII

THE RELATION BETWEEN TEST FACTORS AND PERSON FACTORS

1. *Burt's example, centred both by rows and by columns.*—In the examples we have just considered, there is no doubt that correlations between persons can be calculated without absurdity. In the matrix of marks given by a number of examiners (marking the same paper) to a number of candidates, either two candidates can be correlated or two examiners. The heterogeneity of marks referred to in Chapter XVI, Section 1, does not enter as a difficulty. Still keeping to such material, let us ask ourselves what the relation is between factors found in the one way, and factors found in the other. Qualitatively, we have already suggested that factors and loadings change rôles in some manner. The most determined attempt to find an exact relationship has been that made by Cyril Burt, who concludes that, if the initial units have been suitably chosen, the factors of the one kind of analysis are identical with the loadings of the other, and vice versa (Burt, 1937*b*). The present writer, while agreeing that this is so in the very special circumstances assumed by Burt, is of opinion that his is a very narrow case, and that the factors considered by Burt are not typical of those in actual use in experimental psychology. Theoretically, however, Burt's paper is of very great interest. It can be presented to the general reader best by using Burt's own small numerical example, based on a matrix of marks for four persons in three tests :

	Persons	*a*	*b*	*c*	*d*
	1	−6	2	0	4
Tests	2	3	1	−1	−3
	3	3	−3	1	−1

It will be noticed that this matrix of marks is already centred both ways. The rows add up to zero, and so do the columns. The test scores have been measured from their means, and then thereafter the columns of personal scores have been measured from their means; or it can be done persons first, tests second, the end-result being the same. Burt does not give the matrix of raw scores from which the above matrix comes.

If we take the doubly centred matrix as he gives it, the matrices of variances and covariances formed from it are:

Test Covariances

	1	2	3
1	56	− 28	− 28
2	− 28	20	8
3	− 28	8	20

Person Covariances

	a	*b*	*c*	*d*
a	54	− 18	0	− 36
b	− 18	14	− 4	8
c	0	− 4	2	2
d	− 36	8	2	26

Notice that in both these matrices the columns add to zero, just as they do in the matrices of residues in the "centroid" process.

2. *Analysis of the covariances.*—Burt next proceeds to analyse each of these by Hotelling's method. It seems clear that there will exist some relation between the two analyses, since the primary origin of each matrix is the same table of raw marks, and to show that relation most clearly Burt *analyses the covariances direct*, and not the correlations which could be made from each table (by dividing each covariance by the square root of the product of the two variances concerned). For the two Hotelling analyses he obtains (and the centroid factors before rotation would here be the same):

Analysis of the Tests

$$x_1 = 2\sqrt{14}\gamma_1$$
$$x_2 = -\sqrt{14}\gamma_1 + \sqrt{6}\gamma_2$$
$$x_3 = -\sqrt{14}\gamma_1 - \sqrt{6}\gamma_2$$

Analysis of the Persons

$$a = -3\sqrt{6}f_1$$
$$b = \sqrt{6}f_1 + 2\sqrt{2}f_2$$
$$c = -\sqrt{2}f_2$$
$$d = 2\sqrt{6}f_1 - \sqrt{2}f_2$$

In both cases *two* factors are sufficient (there will always be fewer Hotelling or centroid factors than tests with a doubly centred matrix of marks, for a mathematical reason). The reader can check that the inner products give the covariances, e.g.—

$$\text{covariance}\,(bd) = \sqrt{6} \times 2\sqrt{6} - 2\sqrt{2} \times \sqrt{2} = 12 - 4 = 8$$

The method of finding Hotelling loadings was described in Chapter VII, and the reader can readily check that the coefficients of γ_1, for example, do act as required by that method. For if we use numbers proportional to $2\sqrt{14}$, $-\sqrt{14}$, and $-\sqrt{14}$, namely $1, -\frac{1}{2}, -\frac{1}{2}$, as Hotelling multipliers we get :

56	− 28	− 28	1
− 28	20	8	$-\frac{1}{2}$
− 28	8	20	$-\frac{1}{2}$
56	− 28	− 28	
14	− 10	− 4	
14	− 4	− 10	
84	− 42	− 42	

proportional to $1 \quad -\frac{1}{2} \quad -\frac{1}{2}$ as required.

The largest total (84) is the first "latent root," and the multipliers $1, -\frac{1}{2}, -\frac{1}{2}$, have to be divided, according to Chapter VII, by the square root of the sum of their squares, and multiplied by the square root of 84, giving—

$$2\sqrt{14} \quad -\sqrt{14} \quad -\sqrt{14}$$

3. *Factors possessed by each person and by each test.*—Burt then goes on to "estimate," by "regression equations," the amount of the factors γ possessed by the persons, and the amount of the factors f possessed by the tests. There is a misuse of terms here, for with these factors there is no need to "estimate"; they can be accurately calculated: but that is a small point. The first three equations can be solved for the γ's—there is indeed one equation too many, but it is consistent. And the four equations of the second group can be solved for the f's—again they are consistent. Since the equations are consistent, we can choose the easiest pair in each case to solve for the two unknowns. Choosing the two equations for x_1 and x_2 we obtain—

$$\gamma_1 = \frac{1}{2\sqrt{14}}x_1$$

$$\gamma_2 = \frac{x_2 + \frac{1}{2}x_1}{\sqrt{6}}$$

For the other set of factors we naturally choose the equations in a and c, and have—

$$f_1 = -\frac{a}{3\sqrt{6}}$$

$$f_2 = -\frac{c}{\sqrt{2}}$$

Now, since we are very liable to confusion in this discussion, let us remind ourselves what these factors γ and these factors f are. The factors γ are factors into which each test has been analysed. They do not vary in amount from test to test, but each test is differently loaded with them. They vary in amount from person to person.

The factors f are factors into which each person has been analysed. These do not vary in amount from person to person, but from test to test. Each person is differently loaded with them, that is, made up of them in different proportions. The γ's are uncorrelated fictitious tests: the f's are uncorrelated fictitious persons.

Now, from the equations—

$$\gamma_1 = \frac{1}{2\sqrt{14}}x_1$$

$$\gamma_2 = \frac{x_2 + \frac{1}{2}x_1}{\sqrt{6}}$$

we can find the amount of each factor γ_1 and γ_2 possessed by each person, by inserting his scores x_1 and x_2 in these equations, scores which are given in the matrix :

	a	*b*	*c*	*d*
1	− 6	2	0	4
2	3	1	− 1	− 3
3	3	− 3	1	− 1

Thus the first person possesses γ_1 in an amount $-6/2\sqrt{14}$, because his x_1 is − 6. For the four persons and the two factors we find the amounts of these factors possessed by each person to be :

Factors	γ_1	γ_2
a	$-\frac{3}{\sqrt{14}}$	0
b	$\frac{1}{\sqrt{14}}$	$\frac{2}{\sqrt{6}}$
c	0	$-\frac{1}{\sqrt{6}}$
d	$\frac{2}{\sqrt{14}}$	$-\frac{1}{\sqrt{6}}$

4. *Reciprocity of loadings and factors.*—These are the *amounts* of the factors γ possessed by the four persons. If now the reader will compare them with the *loadings* of the factors f in the second set of equations on page 265, he will see a resemblance. The signs are the same, and the zeros are in the same places. Moreover, the resemblance becomes identity if we destandardize the factors f_1 and f_2, measuring the former in units $\sqrt{84}$ times as large, and the latter in units $\sqrt{12}$ times as large, 84 and 12 being the non-zero latent roots of both matrices. In these units let us

use φ_1 and φ_2 for them. The equations on page 265 giving the analysis of the persons then become—

$$a = \frac{-3\sqrt{6}}{\sqrt{84}}(\sqrt{84}f_1) \qquad\qquad = -\frac{3}{\sqrt{14}}\varphi_1$$

$$b = \frac{\sqrt{6}}{\sqrt{84}}(\sqrt{84}f_1) + \frac{2\sqrt{2}}{\sqrt{12}}(\sqrt{12}f_2) = \frac{1}{\sqrt{14}}\varphi_1 + \frac{2}{\sqrt{6}}\varphi_2$$

$$c = \qquad -\frac{\sqrt{2}}{\sqrt{12}}(\sqrt{12}f_2) = \qquad -\frac{1}{\sqrt{6}}\varphi_2$$

$$d = \frac{2\sqrt{6}}{\sqrt{84}}(\sqrt{84}f_1) - \frac{\sqrt{2}}{\sqrt{12}}(\sqrt{12}f_2) = \frac{2}{\sqrt{14}}\varphi_1 - \frac{1}{\sqrt{6}}\varphi_2$$

It will be seen that the loadings of φ_1 and φ_2 are identical with the amounts of γ_1 and γ_2 in the table on page 267. A similar calculation could be made comparing the amounts of f_1 and f_2 possessed by the tests with the loadings of γ_1 and γ_2 (suitably destandardized) in the analysis of the tests. As we said at the outset, if suitable units are chosen for the marks and the factors, the *loadings* of the personal equations are the *factors* of the test equations, and the *factors* of the personal equations are the *loadings* of the test equations. *But only for doubly centred matrices of marks.* It would be wrong to conclude in general that loadings and factors are reciprocal in persons and tests.

Indeed, even for doubly centred matrices of marks, this simple reciprocity holds only for the analysis of the covariances and not for analyses of the matrices of correlations. Except by pure accident (and as it happens, Burt's example is in the case of test correlations such an accident), the saturations of the correlation analysis will not be any *simple* function of the loadings of the covariance analysis.

5. *Special features of a doubly centred matrix.*—But in any case, a matrix of marks which has been centred both ways is one in which only a very special kind of residual association between the variables is present. Most of what we commonly call the association or resemblance between either tests or persons, the amount of which we gauge by the correlation coefficient, is due to something over and above this. We can write down an infinity of possibly raw

matrices from which Burt's doubly centred matrix might have come. To the rows of the latter matrix we can add *any quantities we like* without in the slightest altering the correlations between the tests, but making enormous changes in the correlations between the persons. Let us, for example, add 10 to the top row, 13 to the middle row, and 16 to the bottom row. There results the matrix :

	a	*b*	*c*	*d*	
1	4	12	10	14	
2	16	14	12	10	(*A*)
3	19	13	17	15	

This gives as correlations between the persons :

	a	*b*	*c*	*d*
a	1·00	·75	·84	− ·14
b	·75	1·00	·28	− ·76
c	·84	·28	1·00	·42
d	−·14	−·76	·42	1·00

Next, without changing this matrix of correlations between persons in the slightest, we can add *any quantities we like* to the columns of the matrix of marks, and produce an infinity of different matrices of correlations between tests. If, for example, we add 5, 2, 8, and 9 to the four columns, we have a matrix of raw marks :

	a	*b*	*c*	*d*	
1	9	14	18	23	
2	21	16	20	19	(*B*)
3	24	15	25	24	

This has the same correlations between persons, but the correlations between tests are now :

	1	2	3
1	1·00	− ·16	·24
2	− ·16	1·00	·92
3	·24	·92	1·00

Or instead, by adding suitable numbers to the columns and to the rows, we might have arrived at the matrix :

	a	b	c	d	
1	44	48	18	10	
2	63	57	27	13	(C)
3	58	48	24	10	

or equally well at :

	a	b	c	d	
1	35	45	37	43	
2	34	34	26	26	(D)
3	34	30	28	28	

The order of merit of the persons in each test is quite different in each of these matrices. The order of difficulty of the tests for each person is quite different in each. If we consider the ordinary correlation between Tests 1 and 2, we find that it is negative in (*B*), zero in (*D*) and positive in (*C*), yet all of these matrices reduce to Burt's matrix when centred both ways. It is clear that they contain factors of correlation which are absent in the doubly centred matrix.

The averages of the rows and the columns of (*C*) are as follows :

	a	b	c	d	Average
1	44	48	18	10	30
2	63	57	27	13	40
3	58	48	24	10	35
Average	55	51	23	11	

The correlation between two tests is clearly influenced very much by the fact that here the person *a* is so much cleverer than the person *d*. Similarly, the correlation between two persons is influenced by the fact that Test 1 is more difficult than Test 2. As soon as the matrix is centred both ways, all the correlation due to these and similar influences is almost extinguished. Centred by rows, (*C*) becomes :

$$\begin{bmatrix} 14 & 18 & -12 & -20 \\ 23 & 17 & -13 & -27 \\ 23 & 13 & -11 & -25 \end{bmatrix}$$

and all the tests are equally difficult on the average. Centred by columns *as well*, it becomes :

$$\begin{bmatrix} -6 & 2 & 0 & 4 \\ 3 & 1 & -1 & -3 \\ 3 & -3 & 1 & -1 \end{bmatrix}$$

and not only are all the tests equally difficult on the average, but all the persons are equally clever on the average. It is to the covariances still remaining that Burt's theorem about the reciprocity of factors and loadings applies. It does not apply to the full covariances of the matrix centred only one way, in the manner usually meant when we speak of covariances or of correlations.

6. *An actual experiment.*—In Part III of Burt's *The Factors of the Mind* (London, 1940) his principle of reciprocity of tests and persons is seen in an actual illustrative experiment on the distribution of temperamental types.

This experiment was on twelve women students, selected because the temperamental assessments made by various judges on them were more unanimous than in the case of the other students. Each, therefore, was a well-marked temperamental type. They were assessed for the eleven traits seen in the table below. The assessments over each trait were standardized, i.e. measured in such units and from such an origin that their sum was zero and the sum of their squares twelve, the number of persons, so that the group was (artificially) made equal in an average of sociability, sex, etc. The correlations between the traits were then calculated and centroid factors taken out, the first two of which I shall call by the Roman letters u and v. These two are *possessed* in some amount by each of the persons and *required*, in degrees indicated by the saturation coefficients, by each of the traits. These saturation coefficients have been found by analysis of the correlations *between the traits*.

Now according to the reciprocity principle, if we analyse instead the correlations *between the persons*, find factors which we may indicate by Greek letters, and measure the amounts of these *possessed* by the eleven traits, these amounts ought to be the same as the saturation coefficients of the Roman factors u, v, etc.

Burt therefore further standardizes the assessments, by persons this time, and finds the total scores on each trait, which are, by a property of centroid factors (see page 217) proportional to the amounts of a centroid Greek factor *possessed* by the eleven traits ; and the test of the reciprocity hypothesis is to see whether these totals are similar to the saturations of a Roman factor. The figures (from Burt's page 405) are given in the table below :

	Saturations of the Roman factors		*Amounts of the Greek factor*
	u	v	α
Sociability	·671	·508	·587
Sex	·878	·213	·489
Assertiveness	·827	·483	·378
Joy	·951	·233	·297
Anger	·824	·241	·280
Curiosity	·780	— ·268	·001
Fear	·898	— ·159	— ·089
Sorrow	·259	— ·104	— ·337
Tenderness	·564	— ·667	— ·447
Disgust	·830	— ·490	— ·489
Submissiveness	·412	— ·685	— ·525

Clearly the amounts of α do not correspond to the saturations of u ; not should they, for a general factor has already been eliminated by the double standardization. They do, however, agree reasonably well with the saturations of the second Roman factor v, and confirm Burt's prediction that, even in this sample, and with factors which are not exactly principal components, the reciprocity principle would still hold approximately.

PART VI

THE INFLUENCE OF SELECTION

CHAPTER XVIII

THE INFLUENCE OF UNIVARIATE SELECTION ON FACTORIAL ANALYSIS*

1. *Univariate selection.*—All workers with intelligence tests know, or ought to know, that the correlations found between tests, or between tests and outside criteria, depend to a very great extent indeed upon the homogeneity or heterogeneity of the sample in which the correlations were measured. If, to take the usual illustration, we measure the correlation between height and weight in a sample of the population which includes babies, children, and grown-ups, we shall obviously get a very high result. If we confine our measurement to young people in their 'teens, we shall usually get a smaller value for the coefficient of correlation. If we make the group more homogeneous still, taking, say, only boys, and all of the same race and exactly the same age, the correlation of height and weight will be still less.† Through all these changes towards greater homogeneity in age, the standard deviation (or its square, the variance) of height has also been sinking, and the standard deviation of weight also. The formulæ which describe these changes were given in 1902 by Professor Karl Pearson,‡ and when the selection of the persons forming the sample is made on the basis of one quality only, these formulæ can be put into the following very simple form.

Let the standard deviations of (say) four qualities be

* Thomson, 1937 and 1938*b*.

† Greater homogeneity need not *necessarily*, in the mathematical sense, decrease correlation, and occasionally it does not do so in actual psychological experiments. But it almost always does so.

‡ These formulæ are not, as was once thought, only applicable if all distributions are normal (see Lawley, 1943*c*, where the necessary conditions are stated). They have been found by trial to give good results even when the sample has been made by cutting off a tail, or both tails, of the distribution.

in the complete population—we must, of course, in each case define what we mean by the complete population, as for example all living adults who were born in Scotland—given by Σ_1, Σ_2, Σ_3, and Σ_4, and their correlations by R_{12}, R_{13}, etc. Now let a selection of persons be made who are more homogeneous in the first quality—say, in an intelligence test which has been given to them all—so that its standard deviation in the sample is only σ_1, and write—

$$\frac{\sigma_1}{\Sigma_1} = p_1$$

The smaller p_1 is, the more homogeneous the group is in intelligence-test score. If we write—

$$q_1 = \sqrt{(1 - p_1^2)}$$

q_1 will be larger, the greater the shrinkage in intelligence score-scatter from Σ_1 to σ_1. We shall call q_1 the " shrinkage " of the quality No. 1 in the sample.

The other qualities 2, 3, and 4, being correlated with the first, will tend to shrink with it, and their expected shrinkages q_2, q_3, and q_4 can be calculated from the formula—

$$q_i = q_1 R_{1i}$$

For the sort of reason indicated earlier in this paragraph, the correlations of the four qualities—which we are for simplicity in exposition assuming to be *positively* correlated in the whole population—will also alter, according to the formula—

$$r_{ij} = \frac{R_{ij} - q_i q_j}{p_i p_j}$$

2. *Elementary proof.*—This formula can be readily proved, for the case where the average is unchanged, by using our geometrical model of correlation, in which tests or other variables are represented by lines all crossing each other at the " average man," and at angles with one another whose cosines equal the correlation coefficients between the tests (see Chapter VI).

In this perspective figure let *OA*, *OB*, and *OC* be three lines in three-fold space representing three tests. The triangle *ABC* is in a plane at right angles to *OA*. Write—

$$\cos \alpha = \cos BOA = R_{12}$$
$$\cos \beta = \cos COA = R_{13}$$
$$\cos \gamma = \cos BOC = R_{23}$$

Take the distance OA as unity. Each test is standardized, so that its standard deviation is unity. Now let the standard deviation of Test 1 be reduced so that it becomes $p_1 = OD$. This means, in our geometrical model, that the whole three-fold space in which our lines OA, OB, and OC exist is compressed from A towards O, and every line parallel to this is shortened in the same way. The point B moves up to E, and the point C to F. The whole triangle ABC is lifted up, remaining at right angles to the line OA, to a new position DEF. The test lines OB and OC become OE and OF. The angle $\gamma = BOC$ has become the angle $\gamma' = EOF$, and $\cos \gamma'$ represents the new correlation coefficient between Tests 2 and 3. Our object is to find $\cos \gamma'$ in terms of the known quantities α, β, γ, and p. One method is to express BC^2 in terms of the triangle BOC, and EF^2 in terms of the triangle EOF, and equate them, since $BC = EF$.

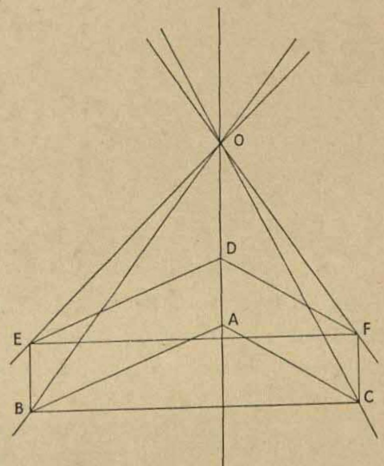


Figure 31.

First note that

$$OB^2 - OE^2 = OA^2 - OD^2 = 1 - p_1^2 = q_1^2$$

and similarly $$OC^2 - OF^2 = q_1^2$$

Also $$p_2 = OE/OB, \text{ and } p_3 = OF/OC$$

Further, $$q_2^2 = 1 - p_2^2 = \frac{OB^2 - OE^2}{OB^2} = q_1^2/OB^2$$

and similarly $$q_3^2 = q_1^2/OC^2$$

Now, since

$$BC^2 = OB^2 + OC^2 - 2OB.OC \cos \gamma$$

and $$EF^2 = OE^2 + OF^2 - 2OE.OF \cos \gamma'$$

we have, subtracting,

$$0=(OB^2-OE^2)+(OC^2-OF^2)-2OB.OC\cos\gamma+2OE.OF\cos\gamma'$$
$$=\quad q_1^2 \quad + \quad q_1^2 \quad -2OB.OC\cos\gamma+2OE.OF\cos\gamma'$$

whence

$$\cos\gamma' = \frac{OB.OC\cos\gamma - q_1^2}{OE.OF}$$

$$= \frac{\cos\gamma - \dfrac{q_1}{OB}\cdot\dfrac{q_1}{OC}}{\dfrac{OE}{OB}\cdot\dfrac{OF}{OC}}$$

$$= \frac{\cos\gamma - q_2q_3}{p_2p_3}$$

or

$$r_{23} = \frac{R_{23} - q_2q_3}{p_2p_3}$$

3. *A numerical example.*—Let us define our "whole population" as all the eleven-year-old children in Massachusetts, and let us suppose (the numbers are entirely fictitious) that the standard deviations of all their scores in four tests are :

1. Stanford-Binet test $16{\cdot}5 = \Sigma_1$,
2. The X reading test $24{\cdot}9 = \Sigma_2$,
3. The Y arithmetic test $27{\cdot}3 = \Sigma_3$,
4. The Z drawing scale $14{\cdot}2 = \Sigma_4$,

while the correlations between these four, in a State-wide survey, are (these are the *R* correlations) :

	1	2	3	4
1	.	·69	·75	·32
2	·69	.	·54	·18
3	·75	·54	.	·06
4	·32	·18	·06	.

Now let a sample of Massachusetts eleven-year-olds be taken who are less widely scattered in intelligence, with a standard deviation in their Stanford-Binet scores of only 10·2. How will all the other quantities listed above tend to alter in this sample ? We have, using the formulæ quoted, the following—

$$p_1 = \frac{10{\cdot}2}{16{\cdot}5} = {\cdot}618$$

$$q_1 = \sqrt{(1 - {\cdot}618^2)} = {\cdot}786$$

and from $q_i = q_1 R_{1i}$ we have the other shrinkages q, and thence the coefficients p and the new standard deviations $\sigma = p\Sigma$:

	1	2	3	4
q	·786	·542	·590	·252
p	·618	·840	·808	·968
σ	10·2	20·9	22·1	13·7

The formula for r_{ij} then enables us at once to calculate the correlations to be expected in the sample, namely :

	1	2	3	4
1	.	·509	·574	·204
2	·509	.	·325	·054
3	·574	·325	.	− ·113
4	·204	·054	− ·113	.

The greater homogeneity in the sample has made all the correlation coefficients smaller, and has indeed made r_{34} become negative.

The reader should note that these standard deviations and correlations are what result from selecting on the Stanford-Binet test, letting the other changes happen in consequence. It would be quite a different matter to select on the X reading test. Even if we did so, so as to reduce the reading test standard deviation from 24·9 to 20·9 as happened above, the other changes would be quite different. The Stanford-Binet standard deviation would, for example, *not* be reduced to 10·2 but only to 15·3. And r_{13} would not be ·574, but ·722. The difference, in terms of our Figure 31, is that whereas selecting the Stanford-Binet corresponded to shortening the line *OA* and with it all parallel distances in the space, selecting the reading test corresponds to shortening OB and all distances parallel to it : quite a different distortion of the space.

4. *From sample to population.*—In the above numerical example we supposed that the standard deviations and correlation coefficients were known in the whole population of Massachusetts eleven-year-old children, and asked what they would become in a sample with a smaller scatter in the Stanford-Binet score. The problem might, however, be reversed, in which case, with a little care, the same formula can be used.

Let us suppose that we know from experiment the above facts about the sample—the standard deviations 10·2, 20·9, 22·1, 13·7, and all the correlation coefficients in the table ·509, ·574, etc.—and that we know further that the standard deviation of Stanford-Binet scores in the whole population in question is 16·5. The sample we have worked with is obviously a biased one, restricted in range of Stanford-Binet scores, and we wish to estimate what our correlation coefficients would have been if we had tested all Massachusetts eleven-year-olds, or, at least, an unbiased sample. We want, indeed, to work the above example backwards.

The quantity p_1 is, in this direction, greater than unity, namely—

$$16{\cdot}5/10{\cdot}2 = 1{\cdot}618$$

and
$$q_1^2 = 1 - p_1^2 = -\ 1{\cdot}617$$

The quantity q_1 is therefore the square root of a minus quantity, which we express as—

$$q_1 = \sqrt{(1{\cdot}617)}i = 1{\cdot}272i, \text{ where } i = \sqrt{-1}$$

The other q's can be got from q_1 by the same formula as before, namely $q_i = q_1R_{1i}$, where R *now* means a correlation coefficient *in the sample.* Thus—

$$q_2 = q_1R_{12} = 1{\cdot}272i \times {\cdot}509 = {\cdot}647i$$
$$q_3 = q_1R_{13} = 1{\cdot}272i \times {\cdot}574 = {\cdot}730i$$

Then—

$$p_2^2 = 1 - q_2^2 = 1 + {\cdot}647^2 \text{ (for } i^2 = -1) = 1{\cdot}419\ ;\ p_2 = 1{\cdot}191$$

and similarly $p_3 = 1{\cdot}238$.

We then have—

$$r_{23} = \frac{R_{23} - q_2 q_3}{p_2 p_3} = \frac{\cdot 325 - \cdot 647i \times \cdot 730i}{1\cdot 191 \times 1\cdot 238}$$

$$= \frac{\cdot 325 + \cdot 472}{1\cdot 475} = \cdot 54$$

as in the table for the population. In this way that table can be completely reconstituted. It is then, of course, only an estimate and, moreover, an estimate based on the assumption that our sample differs from the population only by reason of *one* of the four variables—namely, the Stanford-Binet score—being restricted, deliberately or accidentally, the other restrictions being supposed to have followed sympathetically by reason of the correlations. In few practical examples can we be sure of the mode of selection.

5. *Variance of differences between scores.*—Our numerical example enables us to illustrate a very useful fact, that the variance of the differences between the scores in two tests is independent of the amount of selection if both tests have been equally shrunk, and is reasonably constant when this condition is not too much departed from.

For example, σ^2 for the differences between the scores in Tests 2 and 3 would be, by the formula—

$$\sigma_{2-3}^2 = \sigma_2^2 + \sigma_3^2 - 2r_{23}\sigma_2\sigma_3$$

equal in the population to—

$$24\cdot 9^2 + 27\cdot 3^2 - 2 \times 24\cdot 9 \times 27\cdot 3 \times \cdot 54 = 631\cdot 15$$

and in the sample to—

$$20\cdot 9^2 + 22\cdot 1^2 - 2 \times 20\cdot 9 \times 22\cdot 1 \times \cdot 325 = 625\cdot 0$$

that is, almost the same, although p_2 does not quite equal p_3. This fact gives another method of estimating a population correlation if the sample correlation between differences can be calculated, and if the standard deviations in the population are known or can be guessed. For example, suppose a worker with the sample calculated from his data the value—

$$\sigma_{2-3}^2 = 625$$

and had reason to think that in the population, or in some other sample, the standard deviations were 25 and 27 (as they nearly are in our example), he could estimate the unknown correlation as—

$$\frac{2 \times 25 \times 27 - 625}{2 \times 25 \times 27} = \cdot 537$$

Actually it was ·54. But this method would fail badly if the quantities p_i and p_j were markedly different (Emmett, 1951, *B.J.P.Statist.*, **4**, (1)).

6. *Selection and partial correlation.*—If a sample is made completely homogeneous in the Stanford-Binet test, clearly $p_1 = 0$ and $q_1 = 1$. The same formulæ then give us :

	1	2	3	4
q	1	·69	·75	·32
p	0	·524	·438	·904
σ	0	13·0	11·9	12·8

and the resulting correlation coefficients, which in this case are called " coefficients of partial correlation for *constant* Stanford-Binet score," are, by the same formula :

	1	2	3	4
1	.	.	.	.
2	.	.	·098	— ·086
3	.	·098	.	— ·455
4	.	— ·086	— ·455	.

The correlations of the Stanford-Binet test with the others are given by the formula as 0/0, that is, indeterminate. That they are really zero is seen from the fact that when p_1 is taken as not quite zero, but very small, these correlations come out by the formula as very small. They vanish with p_1.

In this special case of " partial correlation," where the directly selected test is so stringently selected that everyone in the sample has exactly the same score in it, our formula—

$$r_{ij} = \frac{R_{ij} - q_i q_j}{p_i p_j}$$

has a more familiar form. For since—

$$q_i = q_1 R_{1i}$$

and $$q_1 = 1$$

in this case of complete shrinkage we have—

$$q_i = R_{1i}$$

and $$p_i = \sqrt{(1 - R_{1i}^2)}$$

so that our formula becomes—

$$r_{ij} = \frac{R_{ij} - R_{1i}R_{1j}}{\sqrt{(1 - R_{1i}^2)}\ \sqrt{(1 - R_{1j}^2)}}$$

the usual form of a partial correlation coefficient. Its more conventional notation is, calling the test which is made constant Test k instead of Test 1—

$$r_{ij \cdot k} = \frac{r_{ij} - r_{ik}r_{jk}}{\sqrt{(1 - r_{ik}^2)}\ \sqrt{(1 - r_{jk}^2)}}$$

If the " test " which is held constant is the factor g, this becomes—

$$r_{ij \cdot g} = \frac{r_{ij} - r_{ig}r_{jg}}{\sqrt{(1 - r_{ig}^2)}\ \sqrt{(1 - r_{jg}^2)}}$$

which is called the " specific correlation " between i and j. Its numerator is the " residue " left after removing the correlation due to g. If g is the sole cause of correlation, holding g constant will destroy the correlation and we shall have—

$$r_{ij} = r_{ig}r_{jg}$$

as we already saw from another point of view was the case in a hierarchical battery, in Section 4 of Chapter I.

7. *Effect on communalities.*—The formula—

$$r_{ij} = \frac{R_{ij} - q_i q_j}{p_i p_j}$$

is thus a very useful formula, including partial correlation as a special case. If the original variances are each taken as unity, the numerator $R_{ij} - q_i q_j$ for $i \neq j$ gives the new covariances, while p_i^2 and p_j^2 are the new variances.

It also includes as a special case the formula known as the Otis-Kelley formula, which is applicable when two variates have both shrunk to the same extent (a restriction not always recognized). If we put $q_i = q_j$ and therefore $p_i = p_j$ it becomes—

$$p^2 r_{ij} = R_{ij} - q^2 = R_{ij} - 1 + p^2$$
$$p^2(1 - r_{ij}) = 1 - R_{ij}$$
$$\frac{1 - R_{ij}}{1 - r_{ij}} \doteq p^2 = \frac{\sigma_i^2}{\Sigma_i^2} = \frac{\sigma_j^2}{\Sigma_j^2} \text{ the Otis-Kelley formula.}$$

It has a still further application (Thomson, 1938*b*, 456), for if a matrix of correlations in the wider population has been analysed by Thurstone's process, this same formula gives the new communalities (with one exception) to be expected in the sample, if we put $i = j$ and understand by R_{ii}, the communality in the wider population, by r_{ii}, the communality in the sample (and not a reliability coefficient, which is the usual meaning of this symbol). Writing the usual symbol h^2 for communality we have the formula in the form—

$$h_i^2 = \frac{H_i^2 - q_i^2}{p_i^2} \quad (i = 2, 3, 4 \ . \ . \ .)$$

The exception is the new communality of the trait or quality which has been directly selected, in our example No. 1 the Stanford-Binet scores. For the directly selected trait the new communality is given by—

$$h_1^2 = \frac{p_1^2 H_1^2}{1 - q_1^2 H_1^2}$$

(Thomson, 1938*b*, 455; and see also Ledermann, 1938*b*). With these formulæ we can see what is likely to happen to a whole factorial analysis when the persons who are the subjects of the tests are only a sample of the wider population in which the analysis was first made.

8. *Hierarchical numerical example.*—We shall take, in the first place, the perfectly hierarchical example of our Chapter I. But to save space in the tables we shall consider only the first four tests. Their matrix of correlations, with the one common factor and the four specifics added, and with communalities inserted in the diagonal cells, was as follows :

	1	2	3	4	g	s_1	s_2	s_3	s_4
1	(·81)	·72	·63	·54	·90	·44	.	.	.
2	·72	(·64)	·56	·48	·80	.	·60	.	.
3	·63	·56	(·49)	·42	·70	.	.	·71	.
4	·54	·48	·42	(·36)	·60	.	.	.	·80
g	·90	·80	·70	·60	1·00	.	.	.	.
s_1	·44	.	.	.	.	1·00	.	.	.
s_2	.	·60	.	.	.	.	1·00	.	.
s_3	.	.	·71	.	.	.	.	1·00	.
s_4	.	.	.	·80	.	.	.	.	1·00

The bottom right-hand quadrant shows, by its zero entries, that the factors are all uncorrelated with one another, that is, orthogonal. The tests expressed as linear functions of the factors are—

$$z_1 = \cdot 9g + \cdot 436s_1$$
$$z_2 = \cdot 8g + \cdot 600s_2$$
$$z_3 = \cdot 7g + \cdot 714s_3$$
$$z_4 = \cdot 6g + \cdot 800s_4$$

These equations are only another way of expressing the same facts as are shown in the north-east, or the south-west, quadrant of the matrix (where only two places of decimals are used for the specific loadings, to keep the printing regular).

Let us now suppose that this matrix and these equations refer to a wide and defined population, e.g. all Massachusetts eleven-year-olds, and let us ask what will be the most likely matrix of correlations between these tests and factors to be found in a sample chosen by their scores in Test 1 so as to be more homogeneous. The variance of Test 1 in the wider population being taken as unity, let us take that in the more homogeneous select sample as being $p_1^2 = \cdot 36$. We then have, using $q_i = q_1 R_{1i}$, and treating g and the specifics just like tests, the following table :

	1	2	3	4	g	s_1	s_2	s_3	s_4
q	·80	·576	·504	·432	·720	·349	.	.	.
p	·60	·817	·864	·902	·694	·937	1	1	1
p^2 (variance)	·36	·668	·746	·813	·482	·878	1	1	1

For the correlations and communalities, using our formula—

$$\frac{R_{ij} - q_i q_j}{p_i p_j}$$

we get (again printing only two decimal places):

	1	2	3	4	g	s_1	s_2	s_3	s_4
1	(·61)	·53	·44	·36	·78	·28	.	.	.
2	·53	(·46)	·38	·31	·68	−·26	·73	.	.
3	·44	·38	(·32)	·26	·56	−·22	.	·83	.
4	·36	·31	·26	(·21)	·46	−·18	.	.	·89
g	·78	·68	·56	·46	1·00	−·39	.	.	.
s_1	·28	−·26	−·22	−·18	−·39	1·00	.	.	.
s_2	.	·73	.	.	.	.	1·00	.	.
s_3	.	.	·83	.	.	.	.	1·00	.
s_4	.	.	.	·89	.	.	.	.	1·00

In the more homogeneous sample, therefore, the correlations and the communalities of all the tests have sunk. The g column shows what the new correlations of g are with the tests; and on examination of the matrix we see that these, when cross-multiplied with one another, still give the rest of the matrix. Thus—

$$\cdot 78 \times \cdot 46 = \cdot 36 \ (r_{14})$$

$$\cdot 68^2 = \cdot 46 \ (h_2{}^2)$$

The test matrix is still of rank 1 (Thomson, 1938*b*, 453), and these g-column entries can become the diminished loadings of the single common factor required by rank 1.

The columns for the specifics s_2, s_3 (and later specifics also) still show only one entry. In the bottom right-hand quadrant, zero entries show that these specifics are still uncorrelated with one another and with g, that is, g, s_2, s_3, and s_4 are still orthogonal.

But something has happened to the specific s_1. It has become correlated with g, and with all the tests. It has become an oblique factor, orthogonal still to the other specifics, but inclined to g and the tests. It leans further away from Test 1 than it formerly did, and makes obtuse angles (negative correlation) with the other tests and with g, to which it was originally orthogonal.

But since, as we have already pointed out, the test matrix with the reduced communalities is still of rank 1, it is clear that a fresh analysis could be made of the tests into one common factor and specifics, thus—

$$z_1' = \cdot 778g' + \cdot 628s_1'$$
$$z_2' = \cdot 679g' + \cdot 734s_2$$
$$z_3' = \cdot 562g' + \cdot 827s_3$$
$$z_4' = \cdot 462g' + \cdot 887s_4$$

In these equations the factors g', s_1', s_2, s_3, and s_4 are again orthogonal (uncorrelated), and the loadings shown give the correlations and give unit variances. This is the analysis which an experimenter would make who began with the sample and knew nothing about any test measurements in the whole population.

The reader, comparing the loadings in these equations with the correlations in the matrix of the sample, will rightly conclude that the specifics from s_2 onward have not changed. In the matrix it is clear that they are still orthogonal, and their correlations with the tests, in the matrix, are the same as their loadings in the equations. The tests are, in the sample, more heavily loaded with these specifics than they were in the population, but the specifics are the same in themselves.

The new specific s_1' the reader will readily agree to be different from s_1. The latter became oblique in the sample, whereas s_1' is orthogonal. What now is to be said about the common factors g (in the population) and g' (in the sample)? From the fact that the loadings of g', in the sample equations, are identical with the correlations in the sample matrix of the original g with the tests, one is tempted to imagine g' and g to be identical in nature. But that is not so certain.

If we go back to the equations of the tests in the population, we can rewrite them in the following form—

$$z_1 = \cdot 467g' + \cdot 800g'' + \cdot 377s_1'$$
$$z_2 = \cdot 555g' + \cdot 576g'' + \cdot 600s_2$$
$$z_3 = \cdot 485g' + \cdot 504g'' + \cdot 714s_3$$
$$z_4 = \cdot 417g' + \cdot 432g'' + \cdot 800s_4$$

with two common factors g' and g'' instead of one common factor g. These equations still give the same correlations. For example—

$$r_{14} = \cdot467 \times \cdot417 + \cdot800 \times \cdot432 = \cdot540 \text{ as before.}$$

In these equations the specifics s_2, s_3, s_4 are the same, and the communalities of Tests 2, 3, and 4 are the same. All that we have done in these three tests is to divide the common factor g into two components. The ratio of the loading of g'' to the loading of g' is the same in each of them. The loadings of g'' we have made identical with the shrinkages q in the table on page 285.

In Test 1 also we have made the loading of g'' equal to the shrinkage $q_1 = \cdot8$. But in this test g'' cannot be looked upon merely as a component of g. To give the correct correlations, the loading of g' has to be ·467 as shown, and the communality of Test 1 has been raised from its former value (·81) to—

$$\cdot467^2 + \cdot800^2 = \cdot858$$

while the loading of the specific has correspondingly sunk. The factors g', g'', and s_1' are a totally new analysis of Test 1 in the population. Part of the former specific has been incorporated in the common factors.

Now let the factor g'' be abolished, i.e. held constant, so that the tests (now of less than unit variance, so we write them with x instead of z) are—

		Variances
$x_1 = \cdot467g'$	$+ \cdot377s_1'$	·360
$x_2 = \cdot555g'$	$+ \cdot600s_2$	·668
$x_3 = \cdot485g'$	$+ \cdot714s_3$	·746
$x_4 = \cdot417g'$	$+ \cdot800s_4$	·813

The reduced variances are the sum of the squares of the surviving loadings, e.g.—

$$\cdot467^2 + \cdot377^2 = \cdot360$$

The variances, it will be seen, are the p^2's of our tests as measured in the sample. If each of the last set of equations is divided through by the square root of its variance, we arrive at the equations—

$$z_1' = \cdot 778g' + \cdot 628s_1'$$
$$z_2' = \cdot 679g' + \cdot 734s_2$$
$$z_3' = \cdot 562g' + \cdot 827s_3$$
$$z_4' = \cdot 462g' + \cdot 887s_4$$

which is the analysis already given as that of an experimenter who knew only the sample. As to the nature of g', we can say in Tests 2, 3, and 4 that it is possible to regard it as a component of the g of the population. But we cannot do so with assurance in Test 1. There its nature is more dubious. At all events, it is not the same common factor as in the population, and at best we can say that it is one of its components.

9. *A sample all alike in Test* 1.—These phenomena are still more striking if we consider a case where the sample is composed of persons who are all *alike* in Test 1. It would be an excellent exercise for the reader to calculate the resulting matrix of correlations for tests and population factors in this case. The tests act in this case as though their original equations in the population had been—

$$z_1 = g''$$
$$z_2 = \cdot 349g' + \cdot 720g'' + \cdot 600s_2$$
$$z_3 = \cdot 305g' + \cdot 630g'' + \cdot 714s_3$$
$$z_4 = \cdot 262g' + \cdot 540g'' + \cdot 800s_4$$

and then g'' had become zero, i.e. a constant with no variance.

It perhaps helps to a further understanding of what is happening to the factors during selection if we realize that holding the score of Test 1 constant does not hold its factors g and s_1 constant. They can vary in the sample from man to man, but since—

$$z_1 = \cdot 9g + \cdot 436s_1$$

remains constant, a man in the sample who has a high g must have a low s_1—that is, these factors are negatively correlated in the sample. And because they are thus negatively correlated, those members of the sample who have high g's, and who will therefore tend to do well in Tests 2, 3, and 4, will tend to have values below average (negative values) for their s_1, which will be therefore negatively correlated with these tests, in this sample.

So far in our examples we have assumed the sample to be more homogeneous than the population. But a sample can be selected to be less homogeneous. In such a case the same formulæ will serve, if we simply make the capital letters refer to the sample and the small to the population. In fact, the same tables, with their rôles reversed, can illustrate this case. In practical life we usually know which of two groups we would call the sample, and which the population. But mathematically there is no distinction, the one is a distortion of the other, and which is the "true" state of affairs is a question without meaning.

It must also throughout be remembered that all these formulæ and statements refer, not to consequences which are certain to follow, but to consequences which are to be expected. If actual samples were made the values experimentally found in them for correlations, communalities, loadings, etc., would oscillate about those given by our formulæ, violently in the case of small samples, only slightly in the case of large samples.

10. *An example of rank* 2.—The above example has only one common factor. We turn next to consider an example with two. Again it is, we suppose, the first test according to which the sample is deliberately selected, and again we suppose the "shrinkage" q_1 to be ·8. The matrices of correlations and communalities, in the population and in the sample, are then as follows, the two factors f_1 and f_2 and the specifics being treated in the calculation exactly as if they were tests. To economize room on the page, we omit the later specifics :

Correlations in the Population

	1	2	3	4	5	f_1	f_2	s_1	s_2
1	(·65)	·46	·59	·36	·41	·70	·40	·59	.
2	·46	(·37)	·36	·26	·23	·60	·10	.	·79
3	·59	·36	(·61)	·32	·45	·50	·60	.	.
4	·36	·26	·32	(·20)	·22	·40	·20	.	.
5	·41	·23	·45	·22	(·34)	·30	·50	.	.
f_1	·70	·60	·50	·40	·30	(1·00)	.	.	.
f_2	·40	·10	·60	·20	·50	.	(1·00)	.	.
s_1	·59	.	.	.	.	.	.	(1·00)	.
s_2	.	·79	.	.	.	.	.	.	(1·00)

Correlations in the Sample

	1	2	3	4	5	f_1	f_2	s_1	s_2
1	(·40)	·30	·40	·23	·26	·51	·25	·40	.
2	·30	(·27)	·23	·17	·12	·51	−·02	−·21	·85
3	·40	·23	(·50)	·22	·35	·32	·54	−·29	.
4	·23	·17	·22	(·13)	·14	·30	·12	−·16	.
5	·26	·12	·35	·14	(·26)	·15	·44	−·19	.
f_1	·51	·51	·32	·30	·15	(1·00)	−·23	−·36	.
f_2	·25	−·02	·54	·12	·44	−·23	(1·00)	−·18	.
s_1	·40	−·21	−·29	−·16	−·19	−·36	−·18	(1·00)	.
s_2	.	·85	.	.	.	.	.	.	(1·00)

We see here a new phenomenon. The two common factors f_1 and f_2 in the population were orthogonal to one another, as is shown by the zero correlation between them. But in the sample they are negatively correlated (− ·228); that is, they are oblique. We begin to see a generalization which can be algebraically proved, that *all the factors, common and specific, which are concerned with the directly selected test(s) become oblique to each other and to all the tests, but the specifics of the indirectly selected tests remain orthogonal* to everything, except each to its own test.

But *the matrix of the tests themselves is still of rank* 2, and an experimenter working only with the sample would find this out, although he would know nothing about the population matrix. He would therefore set to work to analyse it into two common factors, orthogonal to one another. A Thurstone analysis comes out in two common factors exactly, and can be rotated until all the loadings are positive. For example:

Test	1	2	3	4	5
Factor f_1'	·570	·521	·436	·332	·238
Factor f_2'	·276	.	·555	·130	·452

These factors f', however, are clearly a different pair from the factors f in the original population. In the sample, those original factors (f) are oblique; these (f') are orthogonal.

Again the whole phenomenon is reversible. The second matrix (with the orthogonal factors f') might refer to the population, and a sample picked with a suitable *increased* scatter of Variate 1. All our formulæ could be worked backwards, and we should arrive at the matrix beginning (·65), referring now to the sample. The f' factors would have become oblique, and a new analysis, suitably rotated, would give us the other factors f.

It becomes evident that the orthogonal factors we obtain by the analysis of tests depend upon the subpopulation we have tested. They are not realities in any physical sense of the word; they vary and change as we pass from one body of men to another. It is possible, and this is a hope hinted at in Thurstone's book *The Vectors of Mind*, that if we could somehow identify a set of factors throughout all their changes from sample to sample (in most of which they would be oblique) as being in some way unique, we might arrive at factors having some measure of reality and fixity. Thurstone, in his latest book *Multiple Factor Analysis*, believes that he has achieved this, and that his oblique Simple Structure is invariant. His claim is considered in our next chapter. It is, in the present writer's opinion, justifiable only for univariate selection, not for multivariate, which is not merely repeated univariate selection.

11. *Random selection.*—These considerations deal with the results to be expected when a sample is deliberately selected so that the variance of one test is changed to some desired extent. The new variances and the changed correlations of the other tests given by our formula—

$$r_{ij} = \frac{R_{ij} - q_i q_j}{p_i p_j}$$

are not the *certain* result of our action in selecting for Test 1. If we selected a large number of samples of the same size, all with the same reduced variance in Test 1, they would not all be alike in the resulting correlations. On the contrary, they would all be different. But most of them would be *like* the expected set, few would depart widely from that; and the departures would be in both directions, some samples lying on the one side, others on the other side, of our expectation.

If now, instead of selecting samples which are all alike in the variance of one nominated test, we take a large number of *random* samples of the same size, what would we find ? Among them would be a number which were alike in the variance of Test 1, and these in the other part of the correlation matrix would have values which varied round about those given by our formula. We could also pick out, instead of a set all alike in the variance of Test 1, a different set all alike in the variance of Test 4, say ; and these would have values in the remainder of the matrix oscillating about our formula, in which Test 4 would replace Test 1. In short, a complex family of random samples would show a *structure* among themselves such that if we fix any one variance the average of that array of samples obeys our formula.* Random sampling will not merely add an " error specific " to existing factors, it will make complex changes in the common factors.

* On the author's suggestion, Dr. W. Ledermann has since proved this conjecture analytically (*Biometrika*, 1939*a*, **30**, 295–304). His results cover also the case of multivariate selection (see next chapter).

CHAPTER XIX

THE INFLUENCE OF MULTIVARIATE SELECTION *

1. *Altering two variances and the covariance.*—In the preceding chapter we have discussed the changes which occur in the variances and correlations of a set of tests, and in their factors, when the sample of persons tested is chosen according to their performance in one of the tests : we are next going to see the results of picking our sample by their performances in more than one of the tests, first of all in two of them. Take again, the perfectly hierarchical example of the last chapter. We must this time go as far as six tests in order to see all the consequences. The matrix of correlations of these tests and their factors will be simply an extension of that printed on page 285.

Now let us imagine a sample picked so that the variance of Test 1 and also that of Test 2 is intentionally altered, and further, their covariance (and hence their correlation) changed to some predetermined value.

It is at once clear that in these two directly selected tests the factorial composition will in general be changed —can indeed be changed to anything which is not incompatible with common sense and the laws of logic. What, however, will be the resulting sympathetic changes in the variances and covariances of the other tests of the battery ?

In Chapter XVIII we altered the variance of Test 1 from unity to ·36. The *consequent* diminution in variance to be expected in Test 2 was, as is shown on page 285, from unity to ·668, and the *consequent* change in correlation from ·72 to ·53. Here, however, let us pick our sample so that the variance of the second test is also diminished to ·36, and so that the correlation between them, instead of falling, *rises* to ·833. We have, that is to say, chosen people for our sample who tend to be rather more alike

* Thomson, 1937 ; Thomson and Ledermann, 1938.

than usual in these two test scores, as well as being closely grouped in each, an unusual but not an inconceivable sample. Natural selection (which includes selection by the other sex in mating) has no doubt often preferred individuals in whom two organs tended to go together, as long legs with long arms, and the same sort of thing might occur in mental traits. In terms of variance and covariance we have changed the matrix :

	1	2
1	1·00	·72
2	·72	1·00

$= R_{pp}$

to the matrix :

	1	2
1	·36	·30
2	·30	·36

$= V_{pp}$

for $\frac{\cdot 30}{\sqrt{(\cdot 36 \times \cdot 36)}} = \frac{5}{6} = \cdot 833$, the new correlation. Notice that the diagonal entries here (unities in R_{pp} and ·36, ·36 in V_{pp}) are the variances, not the communalities.

2. *Aitken's multivariate selection formula.*—We shall symbolically represent the whole original matrix of variances and covariances by :

R_{pp}	R_{pq}
R_{qp}	R_{qq}

where the subscript p refers to the directly selected or picked tests, and the subscript q to all the other tests and the factors. R_{pq} (and also R_{qp}) means the matrix of covariances of the picked tests with all the others, including the factors. R_{qq} means the matrix of variances and covariances of the latter among themselves. Since at the outset the tests and factors are all assumed to be standardized, the variances in this whole R matrix are all unity, and the covariances are simply coefficients of correlation. In our case the R matrix is :

Analysis in the Population

	1	2	3	4	5	6	g	s_1	s_2	s_3	s_4	s_5	s_6
1	1·00	·72	·63	·54	·45	·36	·90	·44	.	.	.	.	.
2	·72	1·00	·56	·48	·40	·32	·80	.	·60	.	.	.	.
3	·63	·56	1·00	·42	·35	·28	·70	.	.	·71	.	.	.
4	·54	·48	·42	1·00	·30	·24	·60	.	.	.	·80	.	.
5	·45	·40	·35	·30	1·00	·20	·50	.	.	.	.	·87	.
6	·36	·32	·28	·24	·20	1·00	·40	.	.	.	.	.	·92
g	·90	·80	·70	·60	·50	·40	1·00	.	.	.	.	.	.
s_1	·44	.	.	.	.	.	.	1·00	.	.	.	.	.
s_2	.	·60	.	.	.	.	.	.	1·00	.	.	.	.
s_3	.	.	·71	.	.	.	.	.	.	1·00	.	.	.
s_4	.	.	.	·80	.	.	.	.	.	.	1·00	.	.
s_5	.	.	.	.	·87	.	.	.	.	.	.	1·00	.
s_6	.	.	.	.	.	·92	.	.	.	.	.	.	1·00

The R_{pp} matrix is the square 2×2 matrix, the R_{qq} matrix the square 11×11 matrix, while R_{pq} has two rows and eleven columns, R_{qp} being the same transposed.

Our object is to find what may be expected to happen to the rest of the matrix when R_{pp} is changed to V_{pp}. Formulæ for this purpose were first found by Karl Pearson, and were put into the matrix form in which we are about to quote them by A. C. Aitken (Aitken, 1934). The matrix changes to :

V_{pp}	$V_{pp} R_{pp}^{-1} R_{pq}$
$R_{qp} R_{pp}^{-1} V_{pp}$	$R_{qq} - R_{qp} (R_{pp}^{-1} - R_{pp}^{-1} V_{pp} R_{pp}^{-1}) R_{pq}$

and in order to explain the meaning of these formulæ we shall carry out the calculation for a part of the above matrix only (the first four tests), with a strong recommendation to the reader to perform the whole calculation systematically. If we confine ourselves to the first four tests we have—

$$R_{pp} = \begin{bmatrix} 1{\cdot}00 & {\cdot}72 \\ {\cdot}72 & 1{\cdot}00 \end{bmatrix}$$

$$R_{qq} = \begin{bmatrix} 1{\cdot}00 & {\cdot}42 \\ {\cdot}42 & 1{\cdot}00 \end{bmatrix}$$

$$R_{pq} = \begin{bmatrix} \cdot 63 & \cdot 54 \\ \cdot 56 & \cdot 48 \end{bmatrix}$$

$$R_{qp} = \begin{bmatrix} \cdot 63 & \cdot 56 \\ \cdot 54 & \cdot 48 \end{bmatrix}$$

The most tiresome part of the calculation, if the number of directly selected tests is large, is to find R_{pp}^{-1} the reciprocal of the matrix R_{pp} such that the product—

$$R_{pp} \cdot R_{pp}^{-1} = \begin{bmatrix} 1 & \cdot \\ \cdot & 1 \end{bmatrix} = I$$

where I is the so-called "unit matrix" which has unit entries in the diagonal and zero entries everywhere else.

The method of doing this is given in Chapter XIV, Section 9, page 210. In the present example, where R_{pp} is only of dimensions 2×2, we soon find—

$$R_{pp}^{-1} = \begin{bmatrix} 2\cdot0764 & -1\cdot4950 \\ -1\cdot4950 & 2\cdot0764 \end{bmatrix}$$

When the reciprocal matrix R_{pp}^{-1} has thus been calculated, the best way of proceeding is to find—

$$C = R_{pp}^{-1} R_{pq}$$

and

$$D = R_{qq} - R_{qp} C$$

In the case of our example these are—

$$C = \begin{bmatrix} 2\cdot0764 & -1\cdot4950 \\ -1\cdot4950 & 2\cdot0764 \end{bmatrix} \begin{bmatrix} \cdot 63 & \cdot 54 \\ \cdot 56 & \cdot 48 \end{bmatrix} = \begin{bmatrix} \cdot 4709 & \cdot 4037 \\ \cdot 2209 & \cdot 1894 \end{bmatrix}$$

$$D = \begin{bmatrix} 1\cdot00 & \cdot 42 \\ \cdot 42 & 1\cdot00 \end{bmatrix} - \begin{bmatrix} \cdot 63 & \cdot 56 \\ \cdot 54 & \cdot 48 \end{bmatrix} \begin{bmatrix} \cdot 4709 & \cdot 4037 \\ \cdot 2209 & \cdot 1894 \end{bmatrix}$$

$$= \begin{bmatrix} 1\cdot00 & \cdot 42 \\ \cdot 42 & 1\cdot00 \end{bmatrix} - \begin{bmatrix} \cdot 4204 & \cdot 3604 \\ \cdot 3604 & \cdot 3089 \end{bmatrix}$$

$$= \begin{bmatrix} \cdot 5796 & \cdot 0596 \\ \cdot 0596 & \cdot 6911 \end{bmatrix}$$

subtraction of matrices being carried out by subtracting each element from the corresponding one. We next need—

$$V_{pp} C = \begin{bmatrix} \cdot 36 & \cdot 30 \\ \cdot 30 & \cdot 36 \end{bmatrix} \begin{bmatrix} \cdot 4709 & \cdot 4037 \\ \cdot 2209 & \cdot 1894 \end{bmatrix} = \begin{bmatrix} \cdot 2358 & \cdot 2022 \\ \cdot 2208 & \cdot 1893 \end{bmatrix}$$

which gives us the new covariances of the directly selected tests with those indirectly selected. For V_{qq} we need still $C'(V_{pp}C)$ where the prime indicates that the matrix is transposed (rows becoming columns)—

$$C'(V_{pp}C) = \begin{bmatrix} \cdot 4709 & \cdot 2209 \\ \cdot 4037 & \cdot 1894 \end{bmatrix} \begin{bmatrix} \cdot 2358 & \cdot 2022 \\ \cdot 2208 & \cdot 1893 \end{bmatrix} = \begin{bmatrix} \cdot 1598 & \cdot 1370 \\ \cdot 1370 & \cdot 1175 \end{bmatrix}$$

and then—

$$V_{qq} = D + C'V_{pp}C = \begin{bmatrix} \cdot 5796 & \cdot 0596 \\ \cdot 0596 & \cdot 6911 \end{bmatrix} + \begin{bmatrix} \cdot 1598 & \cdot 1370 \\ \cdot 1370 & \cdot 1175 \end{bmatrix}$$

$$= \begin{bmatrix} \cdot 7394 & \cdot 1966 \\ \cdot 1966 & \cdot 8086 \end{bmatrix}$$

We now can write down the whole new 4×4 matrix of variances and covariances. In the same way, had we included the other tests and the factors, we would have arrived at the whole new 13×13 matrix for all the variances and covariances which we now print.* The values calculated above for the first four tests will be recognized in its top left-hand corner. (The diagonal entries are variances, not communalities.)

Covariances in the Sample

	1	2	3	4	5	6	g	s_1	s_2	s_3	s_4	s_5	s_6
1	·36	·30	·24	·20	·17	·14	·34	·13	·05	.	.	.	.
2	·30	·36	·22	·19	·16	·13	·32	·04	·18	.	.	.	.
3	·24	·22	·74	·20	·16	·13	·33	−·14	−·07	·71	.	.	.
4	·20	·19	·20	·81	·14	·11	·28	−·12	−·06	.	·80	.	.
5	·17	·16	·16	·14	·87	·09	·23	−·10	−·05	.	.	·87	.
6	·14	·13	·13	·11	·09	·92	·19	−·08	−·04	.	.	.	·92
g	·34	·32	·33	·28	·23	·19	·47	−·19	−·10	.	.	.	.
s_1	·13	·04	−·14	−·12	−·10	−·08	−·19	·70	·32	.	.	.	.
s_2	·05	·18	−·07	−·06	−·05	−·04	−·10	·32	·43	.	.	.	.
s_3	.	.	·71	.	.	.	.	.	.	1·00	.	.	.
s_4	.	.	.	·80	.	.	.	.	.	.	1·00	.	.
s_5	.	.	.	.	·87	.	.	.	.	.	.	1·00	.
s_6	.	.	.	.	.	·92	.	.	.	.	.	.	1·00

3. *Features of the sample covariances.*—Examination of this matrix shows the following features :

(1) The specifics of the indirectly selected tests have remained unchanged. They are still orthogonal to each other and all the other tests and factors (except each to

* In such calculations on a larger scale, the methods of Aitken's (1937*a*) paper are extremely economical. Triple products of matrices of the form $XY^{-1}Z$ can thus be obtained in one pivotal operation (see Appendix, paragraph 12).

its own test), are still of unit variance, and have still the same covariances with their own tests, though these will become larger *correlations* when the tests are restandardized ;

(2) The specifics of the directly selected tests have become oblique common factors, correlated with everything except the other specifics ;

(3) The matrix of the indirectly selected tests is still of the same rank (here rank 1) ;

(4) The variances of the factors g, s_1, and s_2 have been reduced to ·47, ·70, and ·43.

An experimenter beginning with this sample, and knowing nothing about the factors in the wider population, would have no means of knowing these relative variances, and would no doubt standardize all his tests. He certainly would not think of using factors with other than unit variance. And even if he were by a miracle to arrive at an analysis corresponding to the last table, with three oblique general factors, he would reject it (*a*) because of the negative correlations of some of the factors, and (*b*) because he can reach an analysis with only two common factors, and those orthogonal. It is therefore practically certain that he will not reach the population factors, at least as far as the directly selected tests are concerned. His data and his analysis will be as overleaf. The variances are all made unity and the covariances converted into correlations. The analysis into factors is a new one, not derived from the last table.

4. *Appearance of a new factor.*—The most noticeable change in this sample analysis, as compared with the population analysis on page 296, is the appearance of a new " factor " h linking the directly selected tests, a factor which is clearly due entirely to that selection. What degree of reality ought to be attributed to it ? Does it differ from the other factors really, or have they also been produced by selection, even in the population, which is only in its turn a sample chosen by natural selection from past generations ?

Otherwise the analysis is still into one common factor and specifics. The loadings of the common factor are

Analysis in the Sample

	1	2	3	4	5	6	g'	h	s_1'	s_2'	s_3	s_4	s_5	s_6
1	1·00	·83	·46	·38	·30	·24	·82	·45	·35	.	.	.	.	.
2	·83	1·00	·43	·35	·28	·22	·77	·45	.	·46	.	.	.	.
3	·46	·43	1·00	·26	·21	·16	·56	.	.	.	·83	.	.	.
4	·38	·35	·26	1·00	·17	·13	·46	.	.	.	.	·89	.	.
5	·30	·28	·21	·17	1·00	·11	·37	.	.	.	.	.	·93	.
6	·24	·22	·16	·13	·11	1·00	·29	.	.	.	.	.	.	·96
g'	·82	·77	·56	·46	·37	·29	1·00	.	.	.	.	.	.	.
h	·45	·45	.	.	.	.	.	1·00	.	.	.	.	.	.
s_1'	·35	.	.	.	.	.	.	.	1·00	.	.	.	.	.
s_2'	.	·46	.	.	.	.	.	.	.	1·00	.	.	.	.
s_3	.	.	·83	.	.	.	.	.	.	.	1·00	.	.	.
s_4	.	.	.	·89	.	.	.	.	.	.	.	1·00	.	.
s_5	.	.	.	.	·93	.	.	.	.	.	.	.	1·00	.
s_6	.	.	.	.	.	·96	.	.	.	.	.	.	.	1·00

less than they were in the population, and this, as our table of variances and covariances shows, is due to a real diminution in the variance of the common factor. The new common factor g' is a component of the old one.

The loadings of s_1 and s_2 have also sunk, because they have been in part turned into a new common factor. The loadings of the other specifics have risen. But this is entirely because the variance of the tests has sunk due to the shrinkage in g, and is not due to any new specifics being added.

All these considerations make it very doubtful indeed whether any factors, and any loadings of factors, have absolute meaning. They appear to be entirely dependent upon the population in which they are measured, and for their definition there would be required not only a given set of tests and a given technical procedure in analysis, but also a given population of persons.

Professor Thurstone, however, in his new book *Multiple Factor Analysis* (1947) gives what he mildly calls " a less pessimistic interpretation than Godfrey Thomson's of the factorial results of selection."

5. *Identity of simple structure factors after univariate selection.*—In that book, Thurstone discusses in Chapter XIX the effects of selection, and shows by examples that

if a battery of tests yields simple structure with oblique factors (including, of course, the orthogonal case), then after *univariate* selection the same factors (though at new angles with one another) are identified by the new structure, which is still simple.

If, for example, the battery which gives the correlations on our page 152, and yields Figure 26 on page 158, has the standard deviation of Test 2 reduced to one-half, then by the methods described on our pages 296–8 we can calculate that the matrix of correlations and communalities becomes :

	1	2	3	4	5	6
1	·589	·295	− ·044	− ·140	·366	·000
2	·295	·302	·049	·159	·183	·000
3	− ·044	·049	·555	·115	·304	·506
4	− ·140	·159	·115	·371	− ·087	·000
5	·366	·183	·304	− ·087	·439	·322
6	·000	·000	·506	·000	·322	·493

The rank of this matrix is still 3 as it was before selection, and three centroid factors are found to have loadings—

	I	II	III
1	·409	·647	·058
2	·379	·244	− ·315
3	·569	− ·444	·184
4	·160	− ·271	− ·522
5	·585	·174	·257
6	·506	− ·350	·337

When these are " extended " in the manner of our page 157 and a diagram like Figure 26 made, we obtain Figure 32. It is still a triangle, and although its measurements are different, the same tests are found defining each side as before. The corners of the triangle may, with Professor Thurstone, reasonably be claimed to represent the same factors as before selection, although their correlations have changed.

The plane of Figure 32 is not the same as the plane of Figure 26, being at right angles to a different first centroid.

When adjustment is made for this, as Professor Thurstone has presumably done in his chapter (though, I protest, without sufficient explanation), then the directly selected

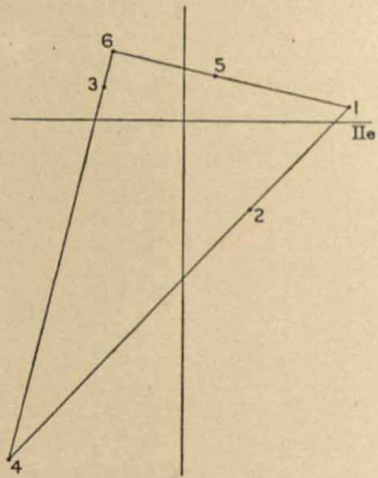


Figure 32.

test point has not moved, while the other points have moved radially away from or towards it.

If the above matrix of centroid loadings is postmultiplied by the rotating matrix obtained from the diagram, viz.—

$$\begin{bmatrix} \cdot721 & \cdot443 & \cdot641 \\ -\cdot499 & -\cdot201 & \cdot744 \\ \cdot480 & -\cdot874 & -\cdot190 \end{bmatrix}$$

we obtain the new simple structure on the reference vectors,

	A	B	C
1	.	.	·732
2	.	·394	·484
3	·562	·180	.
4	.	·472	.
5	·459	.	·455
6	·702	.	.

If this is compared with the table on page 154 it will be seen that the zeros are in the same places, although the non-zero entries have altered (except in Test 6, which was uncorrelated with the directly selected Test 2, and therefore is unaffected in composition).

If the correlations between the factors are calculated by the method of pages 181–2, factor A is found to be still uncorrelated with B and C, but these last two have a correlation coefficient of $-\cdot 3$: that is, they are no longer orthogonal but at an obtuse angle of about $107\frac{1}{2}°$.

6. *Multivariate selection and simple structure.*—But though Thurstone must, I think, be granted his claim that *univariate* selection will not destroy the identity of his oblique simple structure factors, but only change their intercorrelations, the situation would seem to be very different with *multivariate* selection.

Multivariate selection is not the same thing as repeated univariate selection. The latter will not change the rank of the correlation matrix with suitable communalities, nor will it change the position of zero loadings in simple structure. Repeated univariate selection will, it is true, cause all the correlations to alter, but only indirectly and in such a way as to preserve rank, simple structure, and factor identity.

But in multivariate selection it is envisaged that the correlation between two variables may itself be directly selected, and caused to have a value other than that which would naturally follow from the reduction of standard deviation in two selected variables. Selection for correlation is just as easily imagined as is selection for scatter. Indeed, in natural selection it is possibly even commoner.

Once we select for the correlations, however, as well as for scatter, new " factors " emerge, old ones change. In this chapter we have supposed a small part R_{pp} of the whole correlation matrix to be changed to V_{pp}, and found that one new factor is created (page 300) or, indeed, two new *oblique* factors (page 298). We might have supposed R_{pp} to be a larger portion of R : and there is nothing to prevent us supposing selection to go on for the *whole* of R, and writing down a brand-new table of coefficients whose

" factors " would be quite different from those of the original table. In our example of page 152, for instance, where the three oblique " factors " coincided in direction with the communal parts of Tests 1, 4, and 6, there is nothing to prevent us from writing down, as having been produced by selection, a new set of correlation coefficients whose analysis would identify the " factors " with the communal parts of Tests 2, 3, and 5. In fact, all we would have to do would be to renumber the rows and columns on page 152. Such fundamental changes could be produced by selection : and perhaps they have been, for natural selection has had plenty of time at its disposal.

Professor Thurstone (his page 458, footnote, in *Multiple Factor Analysis*) classes the new factors produced by selection as " incidental factors (which) can be classed with the residual factors, which reflect the conditions of particular experiments." But we can hardly dismiss them thus easily if, as is conceivable, they have become the main or perhaps the only factors remaining, the others having disappeared !

It may be admitted at once, however, that the actual amount of selection from psychological experiment to psychological experiment is not likely to make such alarming changes in factors. For the use to which factors are likely to be put in our age, in our century or more, they are like to be independent enough of such selection as can go on in that time, and in that sense Professor Thurstone is justified in his thesis. Nor am I one to deny " reality " to any quality merely because it has been produced by selection, and may not abide for all time.

PART VII

THE NATURE OF FACTORS

CHAPTER XX

THE SAMPLING THEORY

1. *Two views. A hierarchical example as explained by one general factor.*—The advance of the science of factorial analysis of the mind to its present position has not taken place without controversy, and it is the purpose of the present chapter to give a preliminary description of some objections which have been frequently raised by the present writer (Thomson, 1916, 1919*a*, 1935*b*, etc.) which he still holds to.

The contrast between the factorial point of view and Thomson's sampling theory can be best seen by considering the explanation of the same set of correlation coefficients by both views. To simplify the argument we shall take in the first place a set of correlation coefficients whose tetrads are exactly zero, which can therefore be completely "explained" by a general factor *g* and specifics, as in this table :

	1	2	3	4
1	.	·746	·646	·527
2	·746	.	·577	·471
3	·646	·577	.	·408
4	·527	·471	·408	.

We can more exactly follow the argument if we employ the vulgar fractions of which these are the decimal equivalents, namely the following, each divided by 6 :

	1	2	3	4
1	.	$\sqrt{20}$	$\sqrt{15}$	$\sqrt{10}$
2	$\sqrt{20}$	.	$\sqrt{12}$	$\sqrt{8}$
3	$\sqrt{15}$	$\sqrt{12}$	.	$\sqrt{6}$
4	$\sqrt{10}$	$\sqrt{8}$	$\sqrt{6}$	.

In this form the tetrad-differences are all obviously zero by inspection. These correlations can therefore be explained by one general factor, as in Figure 33, which gives them exactly.

We have here a general factor of variance 30 which is the sole cause of the correlations, and specific factors of

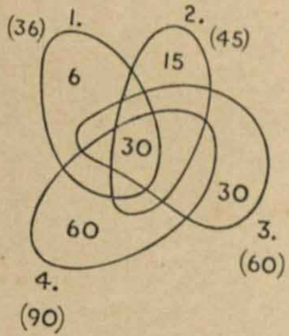


Figure 33.

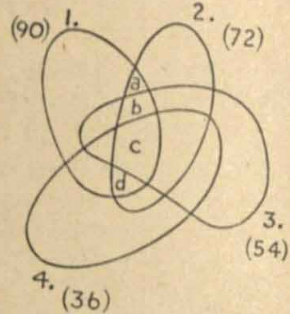


Figure 34.

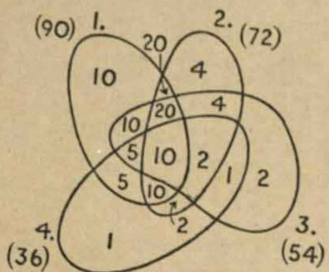


Figure 35.

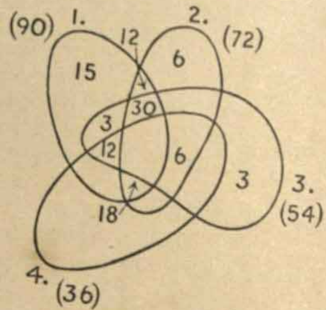


Figure 36.

variances 6, 15, 30, and 60. The variances of the four "tests" are 36, 45, 60, and 90. The "communalities" and "specificities" are :

Test	1	2	3	4	*Totals*
Communality .	$\frac{30}{36}$	$\frac{30}{45}$	$\frac{30}{60}$	$\frac{30}{90}$	$\frac{420}{180} = 2{\cdot}333$
Specificity .	$\frac{6}{36}$	$\frac{15}{45}$	$\frac{30}{60}$	$\frac{60}{90}$	$\frac{300}{180} = 1{\cdot}667$
Totals .	1	1	1	1	4

These communalities can be calculated from the correlation coefficients, for it will be remembered (Chapter I, Section 4) that when tetrad-differences are exactly zero, each correlation coefficient can be expressed as the product of two correlation coefficients with g (two "saturations"). Thus—

$$r_{12} = r_{1g}r_{2g}$$
$$r_{13} = r_{1g}r_{3g}$$
$$r_{23} = r_{2g}r_{3g}$$

Therefore—

$$\frac{r_{12}r_{13}}{r_{23}} = \frac{(r_{1g}r_{2g})\,(r_{1g}r_{3g})}{(r_{2g}r_{3g})} = r_{1g}^2$$

the square of the saturation of Test 1 with g. And when there is only one common factor, the square of its saturation is the communality.

The quantity $r_{12}r_{13}/r_{23}$, therefore, means, on this theory of one common factor, the communality, or square of the saturation with g, of the first test. Its value in our example is 30/36, or five-sixths.

2. *The alternative explanation. The sampling theory.*—The alternative theory to explain the zero tetrad-differences is that each test calls upon a *sample of the bonds* which the mind can form, and that some of these bonds are common to two tests and cause their correlation. In the present instance we have arranged this artificial example so that the tests can be looked upon as samples of a very simple mind, which can form in all 108 bonds (or some multiple of 108).* The first test uses five-sixths of these (or 90), the second test four-sixths (or 72), the third three-sixths (54), and the fourth two-sixths (or 36). These fractions are the same in value as the communalities of the former theory. Each of them may be called the "richness" of the test. Thus Test 1 is most rich, and draws upon five-sixths of the whole mind. The fractions $r_{ij}r_{ik}/r_{jk}$, which in the former theory were "communalities," are in the sampling theory "coefficients of rich-

* There is nothing mysterious about the number 108. It is chosen merely because it leads to no fractions in the diagram. Any large number would do.

ness." They formerly indicated the fraction of each test's variance supplied by g; they indicate here the fraction which each test forms of the whole " mind " (but see later, concerning " sub-pools ").

Now, if our four tests use respectively 90, 72, 54, and 36 of the available bonds of the mind, as indicated in Figure 34, then there may be almost any kind of overlap between two of the tests. Any of the cells of the diagram may have contents, instead of all being empty except for g and the specifics. If we know nothing more about the tests except the fractions we have called their " richnesses," we cannot tell with certainty what the contents of each cell will be; but we can calculate what the *most probable* contents will be. If the first test uses five-sixths and the second test four-sixths of the mind's bonds, it is most probable that there will be a number of bonds common to both tests equal to $\frac{5}{6} \times \frac{4}{6}$, or 20/36ths of the total number. That is, the four cells marked a, b, c, d in the diagram, the cells common to Tests 1 and 2, will most likely contain—

$$\frac{20}{36} \times 108 = 60 \text{ bonds}$$

between them. By an extension of the same principle we can find the most probable number in each cell. Thus c, the number of bonds used in all four of the tests, is most probably—

$$\frac{5}{6} \times \frac{4}{6} \times \frac{3}{6} \times \frac{2}{6} \times 108 = 10 \text{ bonds.}$$

In this way we reach the most probable pattern of overlap of the four tests shown in Figure 35. *And this diagram gives exactly the same correlations as did Figure* 33. Let us try, for example, the value of r_{23} in each diagram. In Figure 33 we had—

$$r_{23} = \frac{30}{\sqrt{(45 \times 60)}} = \frac{\sqrt{12}}{6} = \cdot 577$$

In Figure 35 the same correlation is—

$$r_{23} = \frac{20 + 10 + 4 + 2}{\sqrt{(72 \times 54)}} = \frac{\sqrt{12}}{6} = \cdot 577$$

This form of overlap, therefore, will give zero tetrad-differences, just as the theory of one general factor did. More exactly, this sampling theory gives zero tetrad-differences as the *most probable* (though not the certain) connexion to be found between correlation coefficients (Thomson, 1919*a*) if the sampling of causes is random.

If we let p_1, p_2, p_3, and p_4 represent fractions which the four tests form of the whole pool of N bonds of the mind, then the number common to the first two tests will most probably be p_1p_2N, and the correlation between the tests

$$r_{12} = \frac{p_1p_2N}{\sqrt{(p_1N \cdot p_2N)}} = \sqrt{p_1p_2}$$

We therefore have, in any tetrad, quantities like the following :

	3	4
1	$\sqrt{p_1p_3}$	$\sqrt{p_1p_4}$
2	$\sqrt{p_2p_3}$	$\sqrt{p_2p_4}$

and the tetrad-difference is, most probably (Thomson, 1927*a*, 253)—

$$\sqrt{p_1p_3p_2p_4} - \sqrt{p_1p_4p_2p_3} = 0$$

This may be expressed by saying that the laws of probability alone will cause a tendency to zero tetrad-differences among correlation coefficients. In another form this statement can be worded thus : The laws of probability or chance cause any matrix of correlation coefficients to tend to have rank 1, or at least to tend to have a low rank (where by rank we mean the maximum order among those non-vanishing minors which avoid the principal diagonal elements).

It is, in the opinion of the present writer, this fact—a result of the laws of chance and not of any psychological laws—which has made conceivable the analysis of mental abilities into a few common factors (if not into one only, as Spearman hoped) and specifics. Because of the laws of chance the mind works *as if* it were composed of these hypothetical factors *g*, *v*, *n*, etc., and a number of specific factors. The causes may be "anarchic," meaning that they are numerous and unconnected, yet the result is "monarchic," or at least "oligarchic," in the sense that it may be so described—*provided always that large specific factors are allowed.*

3. *Specific factors maximized.*—The specific factors play, in the usual methods of factorization, an important rôle, and our present example can be used to illustrate the fact, which is not usually realized, that all these methods *maximize the specifics* (Thomson, 1938*c*) by their insistence on *minimizing* the number of common factors. In Figure 33, of the whole variance of 4, the specific factors contribute 1·667, or 41·7 per cent. In Figure 35, they contribute only—

$$\frac{10}{90} + \frac{4}{72} + \frac{2}{54} + \frac{1}{36} = \frac{250}{1{,}080} = \cdot 2315, \text{ or } 5\cdot 8 \text{ per cent.}$$

Apart from certain trivial exceptions which do not occur in practice, it is generally true that minimizing the number of common factors maximizes the variance of the specifics. Innumerable other equivalent analyses of the above correlations can be made, but they all give a variance to the specifics which is less than 1·667. Here, for example, in Figure 36 (page 308), is an analysis which has no general factor but six other common factors, and which gives a total specific variance of—

$$\frac{15}{90} + \frac{6}{72} + \frac{3}{54} + 0 = \frac{330}{1{,}080} = \cdot 3056, \text{ or } 7\cdot 6 \text{ per cent.}$$

Now, specific factors are undoubtedly a difficulty in any analysis, and to have the specific factors made as large and important as possible is a heavy price to pay for having as few common factors as possible.

That specific factors are a difficulty seems to be recognized by Thurstone. "The specific variance of a test," he writes (*Vectors*, 63), "should be regarded as a challenge," and he looks forward to splitting a specific factor up into group factors by brigading the test in question with new companion tests in a new battery. It seems clear that the dissolution of specifics into common factors is unlikely to happen if each analysis is conducted on the principle of making the specific variances as large as possible. We must, however, leave this point here, to return to it later.

4. *Sub-pools of the mind.*—A difficulty which will occur to the reader in connexion with the sampling theory is that, when the correlation between two tests is large, it seems to imply that each needs nearly the whole mind to perform it (Spearman, 1928, 257). In our example the correlation between Tests 1 and 2 was ·746, a correlation not infrequently reached between actual tests. It is, for instance, almost exactly the correlation reported by Alexander between the Stanford-Binet test and the Otis Self-administering test (Alexander, 1935, Table XVI). Does this, then, mean that each of these tests requires the activity of about four-sixths or five-sixths of all the "bonds" of the brain? Not necessarily, even on the sampling theory. These two tests are not so very unlike one another, and may fairly be described as sampling the same region of the mind rather than the whole mind, so that they may well include a rather large proportion of the bonds found in that region. They may be drawn, that is, from a sub-pool of the mind's bonds rather than from the whole pool (Thomson, 1935*b*, 91; Bartlett, 1937*a*, 102). Nor need the phrase "region of the mind" necessarily mean a topographical region, a part of the mind in the same sense as Yorkshire is part of England. It may mean something, by analogy, more like the lowlands of England, all the land easily accessible to everybody, lying below, say, the 300-foot contour line. What the "bonds" of the mind are, we do not know. But they are fairly certainly associated with the neurones or nerve cells of our brains, of which there are probably round about ten thousand million in each normal brain. Thinking is accompanied by the excitation of these neurones in patterns. The simplest patterns are instinctive, more complex ones acquired. Intelligence is possibly associated with the number and complexity of the patterns which the brain can (or could) make. A "region of the mind" in the above paragraph may be the domain of patterns below a certain complexity, as the lowlands of England are below a certain contour line. Intelligence tests do not call upon

brain patterns of a high degree of complexity, for these are always associated with acquired material and with the educational environment, and intelligence tests wish to avoid testing acquirement. It is not difficult to imagine that the items of the Stanford-Binet test call into some sort of activity nearly all the neurones of the brain, though they need not thereby be calling upon all the patterns which those neurones can form. When a teacher is demonstrating to an advanced class that "a quadratic form of rank 2 is identically equal to the product of two linear forms," he is using patterns of a complexity far greater than any used in answering the Binet-Simon items. But the neurones which form these patterns may not be more numerous. Those complicated patterns, however, are forbidden to the intelligence tester, for a very intelligent man may not have the ghost of an idea what a "quadratic form" is. Within the limits of the comparatively simple patterns of the brain which they evoke, it seems very possible that the two tests in question call upon a large proportion of these, and have a large number in common.

As has been indicated, the author is of opinion that the way in which they magnify specific factors is the weak side of the theories of a few common factors. That does not mean, however, that a description of a matrix of correlations in terms of these theories is inexact. Men undoubtedly do perform mental tasks *as if* they were doing so by means of a comparatively small number of group factors of wide extent, and an enormous number of specific factors of very narrow range but of great importance each within its range. Whether a description of their powers in terms of the few common factors only is a good description depends in large measure on what purpose we want the description to subserve. The practical purpose is usually to give vocational or educational advice to the man or to his employers or teachers, and factors, though they cannot improve and indeed may blur the accuracy of vocational estimates, may, however, facilitate them where otherwise they would have been impossible, as money facilitates trade where barter is impossible.

As a theoretical account of each man's mind, however, the theories which use the smallest number of common factors seem to have drawbacks. They can give an exact reproduction of the correlation coefficients. But, because of their large specific factors, they do not enable us to give an exact reproduction of each man's scores in the original tests, so that much information is being lost by their use.

It will be seen from considerations such as these that alternative analyses of a matrix of correlations, even although they may each reproduce the correlation coefficients exactly, may not be equally acceptable on other grounds. The sampling theory, and the single general factor theory, can both describe exactly a hierarchical set of correlation coefficients, and they both give an explanation of why approximately hierarchical sets are found in practice. In a mathematical sense, they are alternatives. But we cannot keep both as realities, though we may employ either mathematically.

5. *The inequality of men.*—Professor Spearman opposed the sampling theory chiefly on the ground that it would make all correlations equal (and zero), and involve the further consequence that all men are equal in their average attainments (*Abilities*, 96), if the number of elementary bonds is large, as the sampling theory requires. Both these objections, however, arise from a misunderstanding of the sampling theory, in which a sample means " some but not all " of the elementary bonds (Thomson, 1935*b*, 72, 76). As has been explained, tests can differ, on this theory, in their richness or complexity, and less rich tests will tend to have low, more complex tests will tend to have high correlations, at any rate if the " bonds " tend to be *all-or-none* in their nature, as the action of neurones is known to be. And as for the assertion that the theory makes all men equal, there is no basis whatever for the suggestion that it assumes every man to have an equal chance of possessing every element or bond. On the contrary, the sampling theory would consider men also to be samples, each man possessing some, but not all, both of the inherited and the acquired neural bonds which are the physical side of thought. Like the tests, some men are

rich, others poor, in these bonds. Some are richly endowed by heredity, some by opportunity and education; some by both, some by neither. The idea that men are samples of all that might be, and that any task samples the powers which an individual man possesses, does not for a moment carry with it the consequences asserted of equal correlations and a humdrum mediocrity among human kind.

6. *Negative and positive correlations.**—The great majority of correlation coefficients reported in both biometric and psychological work are positive. This almost certainly represents an actual fact, namely that desirable qualities in mankind tend to be positively correlated; for though reported correlations may be selected by the unconscious prejudices of experimenters, who are usually on the lookout for things which correlate positively, yet as those who have tried know, it is really very difficult to discover negative correlations between mental tests. Besides, even in imagination we cannot make a race of beings with predominantly negative correlations. A number of lists of the same persons in order of merit can be all very like one another, can indeed all be identical, but they cannot all be the opposite of one another. If Lists a and b are the inverse of one another, List c, if it is negatively correlated with a, will be positively correlated with b. Among a number n of variates, it is logically possible to have a square table of correlation coefficients each equal to unity; that is, an average correlation of unity. But the farthest the average correlation can be pushed in the negative direction is $-1/(n-1)$. That is, if n is large, the average correlation can range from $+1$ to only very little below zero. Even Mother Nature, then, by natural selection or by any other means, could not endow man with abilities which showed both many and large negative correlations. If they were many, they would have to be very small; if they were large, they would have to be very few.

Natural selection has probably tended, on the whole, to

* This section refers to correlations between *tests*. The greater frequency of negative correlations between *persons* has already been discussed in Chapter XVI, Section 8.

favour positive correlations within the species.* In the case of some physical organs it is obvious that a high positive correlation is essential to survival value—for example, between right and left leg, or between legs and arms. In these cases of actual paired organs, however, it is doubtless more than a mere figure of speech to speak of a common factor as the cause. Between organs not simply related to one another, as say eyes and nose, natural selection, if it tended towards negative correlation, would probably split the genus or species into two, one relying mainly on eyesight, the other mainly on smell. Within the one species, since it is mathematically easier to make positive than negative correlations, it seems likely that the former would largely predominate. To say that this was *due to* a general factor would be to hypostatize a very complex and abstract cause. To use a general factor in giving a description of these variates is legitimate enough, but is, of course, nothing more than another way of saying that the correlations are mainly positive—if, as is the case, most

* An important kind of natural selection is the selection of one sex by the other in mating. Dr. Bronson Price (1936) has pointed out that positive cross-correlation in parents will produce positive correlation in the offspring. Price further shows that this positive cross-correlation in the parents will result if the mating is highly homogamous for total or average goodness in the traits, a conclusion which, it may be remarked here, can be easily seen by using the pooling square described in our Chapter XIV. Price concludes: "The intercorrelations which *g* has been presumed to illumine are seen primarily as consequences of the social and therefore marital importance which has attached to the abilities concerned." Price in his argument makes use of formulæ from Sewall Wright (1921). M. S. Bartlett, in a note on Price's paper (Bartlett, 1937*b*), develops his argument more generally, also using Wright's formulæ, and says: "Price contrasts the idea of elementary genetic components with factor theories. . . . It should, however, be pointed out that a statistical interpretation of such current theories can be and has been advocated. Thomson has, for example, shown . . .", and here follows a brief outline of the sampling theory. "On the basis of Thomson's theory," Bartlett adds, "I have pointed out (Bartlett, 1937*a*) that general and specific abilities may naturally be defined in terms of these components, and that while some statistical interpretation of these major factors seems almost inevitable, this may not in itself render their conception invalid or useless."

people mean by a general factor one which *helps* in every case, not an interference factor which sometimes helps and sometimes hinders.

7. *Low reduced rank.*—It is, however, on the tendency to a low reduced rank in matrices of mental correlations that the theory of factors is mainly built. It has very much impressed people to find that mental correlations can be so closely imitated by a fairly small number of common factors. Ignoring the host of large specific factors to which this view commits them, they have concluded that the agreement was so remarkable that there must be something in it. There is ; but it is almost the opposite of what they think. Instead of showing that the mind has a definite structure, being composed of a few factors which work through innumerable specific machines, the low rank shows that the mind has hardly any structure. If the early belief that the reduced rank was in all cases *one* had been confirmed, that would indeed have shown that the mind had no structure at all but was completely undifferentiated. It is the *departures* from rank 1 which indicate structure, and it is a significant fact that a general tendency is noticeable in experimental reports to the effect that batteries do not permit of being explained by as small a number of factors in adults as in children, probably because in adults education and vocation have imposed a structure on the mind which is absent in the young.

By saying that the mind has little structure, nothing derogatory is meant. The mind of man, and his brain, too, are marvellous and wonderful. All that is meant by the absence of structure is the absence of any fixed or strong linkages among the elements (if the word may for a moment be used without implications) of the mind, so that any sample whatever of those elements or components can be assembled in the activity called for by a " test."

Not that there is any necessity to suppose that the mind is composed of separate and atomic elements. It is possibly a continuum, its elements if any being more like the molecules of a dissolved crystalline substance than like grains of sand. The only reason for using the word " elements " is that it is difficult, if not impossible, to speak of the different parts of the mind without assuming some "items" in terms of which to think. For concreteness it is convenient to identify the elements, on the mental side, with something of the nature of Thorndike's "bonds," and on the bodily side with neurone arcs; in the remainder of this chapter the word "bonds" will be used. But there is no necessity beyond that of convenience and vividness in this. The "bonds" spoken of may be identified by different readers with different entities. All a "bond" means, is some very simple aspect of the causal background. Some of them may be inherited, some may be due to education. There is no implication that the combined action of a number of them is the mere sum of their separate actions. There is no commitment to "mental atomism."

If, now, we have a causal background comprising innumerable bonds, and if any measurement we make can be influenced by any sample of that background, one measurement by this sample and another by that, all samples being possible; and if we choose a number of different measurements and find their intercorrelations, the matrix of these intercorrelations will tend to be hierarchical, or at least tend to have a low reduced rank. This has nothing to do with the mind: it is simply a mathematical necessity, whatever the material used to illustrate it.

8. *A mind with only six bonds.*—We shall illustrate this fact first by imagining a "mind" which can form only six "bonds," which mind we submit to four "tests" which are of different degrees of richness, the one requiring the joint action of five bonds, the others of four, three, and two respectively (Thomson, 1927*b*). These four tests will (when we give them to a number of such minds) yield correlations with one another. For we shall suppose the different minds not all to be able to form all six of the possible bonds, some individuals possessing all six, others possessing smaller numbers.

We have only specified the richness of each test, but have not said *which* bonds form each ability. There may, therefore, be different degrees of overlap between them, though some will be more frequent than others if we form all the possible sets of four tests which are of richness five, four, three, and two. If we call the bonds *a*, *b*, *c*, *d*, *e*, and *f*, then one possible pattern of overlap would be the following:

Test	*Bonds*					
1	*a*	*b*	*c*	*d*	*e*	.
2	.	*b*	*c*	*d*	*e*	.
3	.	.	.	*d*	*e*	*f*
4	.	.	*c*	*d*	.	.

If we for further simplicity suppose these bonds to be equally important, and use the formula—

$$\text{Correlation} = \frac{\text{overlap}}{\text{geometrical mean of the two totals}}$$

we can calculate the correlations which these four tests would give, namely:

	1	2	3	4
1	.	$\frac{4}{\sqrt{20}}$	$\frac{2}{\sqrt{15}}$	$\frac{2}{\sqrt{10}}$
2	$\frac{4}{\sqrt{20}}$	.	$\frac{2}{\sqrt{12}}$	$\frac{2}{\sqrt{8}}$
3	$\frac{2}{\sqrt{15}}$	$\frac{2}{\sqrt{12}}$	.	$\frac{1}{\sqrt{6}}$
4	$\frac{2}{\sqrt{10}}$	$\frac{2}{\sqrt{8}}$	$\frac{1}{\sqrt{6}}$	.

and we notice that in this particular pattern all three tetrad-differences are zero. However, if we picked our four tests at random (taking care only that they were of these degrees of richness) we would not always or often get the above pattern: in point of fact, we would get it only 12 times in 450. Nevertheless, it is one of the most probable patterns. In all, 78 different patterns of the bonds are possible—always adhering to our five, four, three, and two—the probability of each pattern ranging from 12 in 450 down to 1 in 450.

It is possible to calculate the tetrad-differences for each one of the 78 possible patterns of overlap which can occur. When we then multiply each pattern by the expected frequency of its occurrence in 450 random choices of the four tests, we get 450 values for each tetrad-difference, distributed as follows :

Values of $F \times \sqrt{120}$	*Frequency of* F_1	F_2	F_3
8			2
7		4	0
6		8	14
5	9	2	6
4	27	34	23
3	6	12	30
2	75	72	48
1	61	66	72
0	99	54	81
− 1	56	78	36
− 2	67	42	42
− 3	16	30	60
− 4	30	36	18
− 5	0	0	0
− 6	4	12	18
	450	450	450

Although the distribution of each F about zero is slightly irregular, *the average value of each F is exactly zero.* For F_1 the variance is—

$$\sigma^2 = \frac{2{,}164}{120 \times 450} = \cdot 040$$

We see, then, that in this universe of very primitive-minded men, whose brains can form only six bonds, four tests which demanded respectively five, four, three, and two bonds would give tetrad-differences whose expected value would be zero, the values actually found being grouped around zero with a certain variance. There is no particular mystery about the four " richnesses " five, four, three, and two, by the way. We might have taken any

" richnesses " and got a similar result. If there are no linkages among the bonds, the most probable value of a tetrad-difference will always be zero ; and if all possible combinations of the bonds are taken, the average of all the tetrad-differences will be zero. With only six bonds in the " mind," however, the scatter on both sides of zero will be considerable, as the above value of the standard deviation of F_1 shows, viz.—

$$\sigma = \sqrt{\cdot 040} = \cdot 20$$

9. *A mind with twelve bonds.*—But as the number of bonds in the mind increases, the tetrad-differences crowd closer and closer to zero. Let us, for example, suppose exactly the same experiment as above conducted in a universe of men whose minds could form twelve bonds (instead of six), the four tests requiring ten, eight, six, and four of these (instead of five, four, three, and two) (Thomson, 1927*b*). This increase in complexity enormously increases the work of calculating all the possible patterns of overlap, and the frequency of each. There are now 1,257 different square tables of correlation coefficients and still more patterns of overlap, some of which, however, give the same correlations. When each possibility is taken in its proper relative frequency (ranging from once to 11,520 times) there are no fewer than 1,078,110 instances required to represent the distribution. They have, nevertheless, all been calculated, and the distribution of F_1 was as follows :

$\sqrt{1920}\ F_1$	*Freq.*	$\sqrt{1920}\ F_1$	*Freq.*	$\sqrt{1920}\ F_1$	*Freq.*	$\sqrt{1920}\ F_1$	*Freq.*
20	225	7	17,760	− 3	31,432	− 13	624
18	1,800	6	74,392	− 4	72,676	− 14	3,792
16	1,755	5	15,744	− 5	53,808	− 15	4,144
15	4,600	4	52,085	− 6	49,328	− 16	3,970
14	3,840	3	121,608	− 7	21,240	− 18	112
12	19,610	2	42,384	− 8	41,951	− 19	456
11	10,632	1	28,096	− 9	5,896	− 20	584
10	8,360	0	122,699	− 10	29,184	− 24	28
9	26,696	− 1	63,024	− 11	8,960		
8	37,735	− 2	81,208	− 12	15,672		

Total 1,078,110

This table again gives an average value of F_1 exactly equal to zero. But the separate values of the tetrad-difference are grouped more closely round zero than before, with a variance now given by—

$$\sigma^2 = \frac{37,166,400}{1,920 \times 1,078,110} = 0{\cdot}018$$

This is rather less than half the previous variance. Doubling the number of bonds in the imagined mind has halved the variance of the tetrad-differences. If we were to increase the number of potential bonds supposed to exist in the mind to anything like what must be its true figure, we would clearly reach a point where the tetrad-differences would be grouped round zero very closely indeed.

The principle illustrated by the above concrete example can be examined by general algebraic means, and the above suggested conclusion fully confirmed (Mackie, 1928*a*, 1929). It is found that the variance of the tetrad-differences sinks in proportion to $1/(N - 1)$, where N is the number of bonds, when N becomes large, and the above example agrees with this even for such small N's as 6 and 12 : for—

$$\frac{6 - 1}{12 - 1} \times {\cdot}040 = {\cdot}018 \text{ as found.}$$

In this mathematical treatment, bonds have been spoken of as though they were separate atoms of the mind, and, moreover, were all equally important. It is probably quite unnecessary to make the former assumption, which may or may not agree with the actual facts of the mind, or of the brain. Suitable mathematical treatment could probably be devised to examine the case where the causal background is, as it were, a continuum, different proportions of it forming tests of different degrees of richness. And as for the second assumption, it is in all likelihood merely formal. Let the continuum be divided into parts of equal importance, and then the number of these increased and their extent reduced, keeping their importance equal. What is necessary, to give the result that zero tetrads are so highly probable, is *that it be possible to take our tests* *with equal ease from any part of the causal background; that there be no linkages among the bonds* which will disturb the random frequency of the various possible combinations; *in other words, that there be no "faculties" in the mind.* And it is also necessary that all possible tests be taken in their probable frequency.

In any actual experiment, of course, it is quite impracticable to take all possible tests, which are indeed infinite in number. A sample of tests is taken. If this sample is large *and random*, then there should, in a mind without separate "faculties," without linkages between its bonds, be an approach to zero tetrads. The fact that this tendency attracted Professor Spearman's attention, and was sufficiently strong to make him at first believe that all samples of tests showed it, provided care was taken to avoid tests so alike as to be almost duplicates (which would be "statistical impossibilities" in a random sample), indicates that the mind is indeed very free to use its bonds in any combination, that they are comparatively unlinked.

The sampling theory assumes that each ability is composed of *some but not all* of the bonds, and that abilities can differ very markedly in their "richness," some needing very many "bonds," some only few. It further requires some approach to "all-or-none" reaction in the "bonds"; that is, it supposes that a bond tends either not to come into the pattern at all, or to do so with its full force. This does not seem a very unnatural assumption to make. It would be fulfilled if a "bond" had a threshold below which it did not act, but above which it did act; and this property is said to characterize neurone arcs and patterns. When this form of sampling is assumed the rank of the correlation matrix tends to be reducible to a small number, *if all possible correlations are taken*, and finally to be *one* as the bonds increase without limit.

It is important to realize what is meant by the rank *tending* to rank 1 as more and more of the possible correlations are taken. When the rank is 1 the tetrad-differences are zero. But clearly, the reader may say, taking more and more samples of the bonds to form more and more tests will not change in any way the pre-existing tetrad-differences, will not make them zero if they are not zero to start with. That is perfectly true; but that is not what is meant. As more and more tests are formed by samples of the bonds, the number of zero and very small tetrads will increase and swamp the large tetrads. The sampling theory does not say that all tetrads will be exactly zero, or the rank exactly 1. It says that the tetrads will be distributed about zero (not because each is taken both plus and minus, but when all are given their sign by the same rule) with a scatter which can be reduced without limit, in the sense that with more bonds the *proportion* of large tetrads becomes smaller and smaller; always provided all possible samples are taken, i.e. that the family of correlation coefficients is complete.

With a finite number of tests this, of course, is not the case; but if the tests are a random sample of all possible tests, there will again be the approach to zero tetrads. The same will be true if the tests are sampling not the whole mind, but some portion of it, some sub-pool of our mind's abilities. If we stray from this pool and fish in other waters, we shall break the hierarchy; but if we sampled the *whole* pool of a mind, we should again find the tendency to hierarchical order. If the mind is *organized* into sub-pools (such as the verbal sub-pool, say), then we shall be liable to fish in two or three of them, and get a rank of 2 or 3 in our matrix, i.e. get two or three common factors, in the language of the other theory.

10. *Contrast with physical measurements.*—The tendency for tetrad-differences to be closely grouped around zero appears to be stronger in mental measurements than elsewhere; stronger, for example, than in physical measurements although it is found there too.

In physical measurements we do not measure a person's body just from anywhere to anywhere. We observe organs and measure them—leg, cranium, chest girth, etc. The variates are not a random sample. In other words, the physical body has an obvious structure which guides our measurements, and the tendency to a low rank among the correlation coefficient, although present, is less than among mental measurements. The tendency to zero tetrad-differences in the mind is due to the fact that the mind has, comparatively speaking, no organs. We can, and do, measure it almost from anywhere to anywhere. No test measures a leg or an arm of the mind ; every test calls upon a group of the mind's bonds which intermingles in most complicated ways with the groups needed for other tests, without being a set pattern immutably linked into an organ. Of all the conceivable combinations of the bonds of the mind we can, without great difficulty, take a random sample, whereas in physical measurements we take only the sample forced on us by the organs of the body. Being free to measure the mind almost from anywhere to anywhere, we can get a set of measurements which show " hierarchical order " without overgreat trouble. We can do so because the mind is so comparatively structureless. Mental measurements tend to show hierarchical order, and to be susceptible of mathematical description in terms of one general factor or few, and innumerable specifics, not because there are specific neural machines through which its energy must show itself, but just exactly because there are no fixed neural machines. The mind is capable of expressing itself in the most plastic and Protean way, especially before education, language, the subjects of the school curriculum, the occupation, and the political beliefs of adult life have imposed a habitual structure on it. It is not without significance that the " factor " most widely recognized after Spearman's *g* is the verbal factor *v*, the mother-tongue being, as it were, the physical body of the mind, its acquired structure.

11. *Absolute variance of different tests.*—It will be noted that on the sampling theory the different tests will naturally have different variances, the " richer " tests having a wider scatter. This seems only natural. It is customary, at any rate in theoretical discussions, to reduce all scores in different tests to standard measure, thereby equalizing their variance. This seems inevitable, for there is no means of comparing the scatter of marks in two different tests. But it does not follow that the scatter would be really the same if some means of comparison were available. When the same test is given to two

different groups we have no hesitation in ascribing a wider variance to the one or the other group, and it seems conceivable that a similar distinction might mentally be made between the scores made by one group in two different tests. The writer is completely in accord with M. S. Bartlett when he says (Bartlett, 1935, 205): "I think many people would agree . . . that the variation in mathematical ability displayed even in a selected group such as Cambridge Tripos candidates cannot be altogether put down to the method of marking adopted by the examiners." We may put these mathematics marks into standard measure, and we may put the marks scored by the same group in, say, a form-board test, also into standard measure. But that does not imply that at bottom the two variances are equal, if only we had some rigorous way of comparing them. Our common sense tells us plainly that they are not equal in the absolute sense, though for many purposes their difference is irrelevant. It seems to be no defect, then, but rather a good quality, of the sampling theory to involve different absolute variances.

12. *A distinction between* g *and other common factors.*—The writer is inclined to make a distinction in interpretation between the Spearman general factor *g* and the various other common factors, mostly if not all of less extent than *g*, which have been suggested. When properly measured by a wide and varied hierarchical battery, *g* appears to him to be an index of the span of the whole mind, other common factors to measure only sub-pools, linkages among bonds. The former measures the whole number of bonds; the latter indicate the degree of structure among them.

Some of this "structure" is no doubt innate; but more of it is probably due to environment and education and life. Its expression in terms of separate uncorrelated factors suggests what is almost certainly not the case, that the "sub-pools" are separate from one another. The actual organization is likely to be much more complicated than that, and its categories to be interlaced and interwoven, like the relationships of men in a community, plumbers and Methodists, blonds, bachelors, smokers, Conservatives, illiterates, native-born, criminals, and school-teachers, an organization into classes which cut across one another right and left.

Further, it is improbable that the organization of each mind is the same. The phrase "factors of *the* mind" suggests too strongly that this is so, and that minds differ only in the amount of each factor they possess. It is more than likely that different minds perform any task or test by different means, and indeed that the same mind does so at different times.

Yet with all the dangers and imperfections which attend it, it is probable that the factor theory will go on, and will serve to advance the science of psychology. For one thing, it is far too interesting to cease to have students and adherents. There is a strong natural desire in mankind to imagine or create, and to name, forces and powers behind the façade of what is observed, nor can any exception be taken to this if the hypotheses which emerge explain the phenomena as far as they go, and are a guide to further inquiry. That the factor theory has been a guide and a spur to many investigators cannot be denied, and it is probably here that it finds its chief justification.

CHAPTER XXI

SOME FUNDAMENTAL QUESTIONS

IT seems advisable to conclude with a brief discussion of some of the fundamental theoretical questions needing an answer. Among these are the following, of which (1) and (3) are rather liable to be forgotten by those actually engaged in making factorial analyses :

(1) What metric or system of units is to be used in factorial analysis ?

(2) On what principle are we to decide where to stop the rotation of our factor-axes or how to choose them so that rotation is unnecessary ?

(3) Is the principle of minimizing the number of common factors, i.e. of analysing only the communal variance, to be retained ?

(4) Are oblique, i.e. correlated factors to be permitted ?

1. *Metric.*—Most of the work done in factorial analysis has assumed the scores of the tests to be standardized ; that is to say, in each test the unit of measure has been the actual standard deviation found in the distribution. This is in a sense a confession of ignorance. The accidental standard deviation which happens to result from the particular form of scoring used in a test means, of course, nothing more. Yet there is undoubtedly something to be said for the probability of real differences of standard deviation existing between tests (see Chapter XX, Section 11). In that case, if we knew these real standard deviations, we would use variances and covariances and analyse them, not correlations (compare Hotelling, 1933, 421–2 and 509–10).

Burt has urged the use of variances and covariances, which are indeed necessary to him to enable his relation between trait factors and person factors to hold (see Chapter XVII, page 264). But the variances and covariances he actually uses are simply the arbitrary ones which arise from the raw scores, and depend entirely upon the scoring system used in each test. It would seem necessary to have some system of rational, not arbitrary, units.

Hotelling has already suggested one such, based upon the idea of the principal components of all possible tests, but it would seem to be unattainable in practice (Hotelling, 1933, 510). Another can be based on the ideas of the sampling theory and has already been foreshadowed in the previous chapter. Tests quite naturally have different variances on that theory, since they comprise larger or smaller samples of the " bonds " of the mind (see Thomson, 1935*b*, 87). In a hierarchical battery these natural variances are measured by the " coefficient of richness." The " richness " of Test k is given by

$$\frac{r_{ik}r_{jk}}{r_{ij}},$$

the same quantity as the square of Spearman's " saturation with g." It is, on the sampling theory, the fraction which the test forms of the pool of bonds which is being sampled, and is the natural variance of the test in comparison with other tests from that pool. The " saturation with g " of Spearman's theory is the " natural standard deviation " of the sampling theory. Even in a battery which is not hierarchical, the formula (Chapter III, Section 5, page 43)—

$$\sqrt{\frac{A^2 - A'}{T - 2A}}$$

will give a rough estimate of the natural standard deviation of each test. The general principle is that tests which show the most total correlation have the largest natural variance.

2. *Rotation.*—Our views on the rotation of factors will depend on what we want them to do. Burt looks upon them as merely a convenient form of classification and is content to take the principal axes of the ellipsoids of density, or that approximation to them given by a good centroid analysis, as they stand, without any rotation. He " takes out " the first centroid factor, either by calculation or by selecting a very special group of persons each of whom

has in a battery of tests an average score equal to the population average, each of the tests also having the same average as every other test in the battery over this subgroup of persons (Burt, 1938*a*). He concentrates attention on the remaining factors, which are "bipolar," having both positive and negative weights in the tests. When, as in the article referred to, he is analysing temperaments, this fits in well with common names for emotional characteristics, for those names too are usually bipolar, as brave-cowardly, extravagant-stingy, extravert-introvert, and so on.

Thurstone, on the other hand, emphatically insists on the need for rotation if the factors are to have psychological meaning (Thurstone, 1938*a*, 90). The centroid factors are mere averages of the tests which happen to form the battery, and change as tests are added or taken away, whereas he wants factors which are invariant from battery to battery. I think he would put invariance before psychological meaning, and say that if a certain factor keeps turning up in battery after battery we must ask ourselves what its psychological meaning is. His own opinion, backed up by a great deal of experimental work of a pioneering and exploratory nature, is that his principle of rotating to "simple structure" gives us also psychologically meaningful and invariant factors.

The problems of rotation and metric are not unconnected, and one piece of evidence in favour of rotating to simple structure is that the latter is independent of the units used in the tests. If instead of analysing correlations we analyse covariances, with whatever standard deviations we care to assign to the tests, we get a centroid analysis quite different from the centroid analysis of correlations. But if we rotate each to simple structure the tables are identical, except, of course, that in the covariance structure each row is multiplied by the standard deviation of the test.

For example, if we take the six tests of Chapter XI Section 2 (page 152) and ascribe arbitrary standard deviations of 1, 2, 3, 4, 5, and 6 to them, we can replace the correlations and communalities by covariances and variance-communalities, and perform a centroid analysis. Since we know the proper communalities* it comes out exactly in three factors with no residues, and gives the centroid structure :

	I	*II*	*III*
1	·372	·567	·462
2	·948	1·278	− ·060
3	1·969	−1·016	− ·337
4	1·002	1·072	−2·118
5	2·992	·593	1·716
6	3·379	−2·493	·337

When this is rotated to simple structure, by post-multiplication by the matrix

$$\begin{bmatrix} \cdot 802 & \cdot 389 & \cdot 453 \\ -\cdot 592 & \cdot 416 & \cdot 691 \\ \cdot 080 & -\cdot 822 & \cdot 564 \end{bmatrix}$$

the resulting table is :

	A	*B*	*C*
1	.	.	·820
2	.	·950	1·278
3	2·154	·619	.
4	.	2·577	.
5	2·187	.	2·732
6	4·213	.	.

This is identical with the simple structure found from the correlations, if the rows here are divided by 1, 2, 3, 4, 5, and 6, the standard deviations. It is definitely a point in favour of simple structure that it is thus independent of the system of units employed. Spearman's analysis of a hierarchical matrix into one *g* and specifics also has this

* If we have to guess communalities, our two simple structures will differ slightly because the highest covariance in a column may not correspond to the highest correlation. But with a battery of many tests this difference will be unimportant, and could be annulled by iteration.

property of independence of the metric. If the tetrad-differences of a matrix of correlations are zero, and we analyse into one general factor and specifics, it is immaterial whether we analyse correlations or covariances. The loadings obtained in the latter case are exactly the same except, of course, that each is multiplied by the appropriate standard deviation.

At this point one is reminded of Lawley's loadings* found by the method of maximum likelihood, for these possess the property that the unrotated loadings obtained from correlations are already the same as the *unrotated* loadings obtained from covariances, if the latter are divided by the standard deviations. Centroid analyses, or principal component analyses, do not possess this property. The loadings obtained by these means from covariances cannot be simply divided by the standard deviations to give the loadings derived from correlations, though the one can be rotated into the other. Lawley's loadings need no such rotation. They are, as it were, at once of the same *shape* whether from covariances or from correlations and only need an adjustment of units, such as one makes in changing, say, from yards to feet. A field which is 50 yards broad and 20 poles long has the same shape as one which is 150 feet broad and 330 feet long.

Now, as we have seen, this property of equivalence of covariance and correlation loadings is also possessed by simple structure. It would thus not be unnatural to hope that Lawley's method might lead straight to simple structure, without any rotation. But this is not the case. Clearly, then, simple structure is not the only position of the axes where the loadings are independent of the units of measurement employed. Indeed, any subsequent post-

* In accordance with our definition on page 170, the term " loading " means a coefficient in a specification equation, an entry in a " pattern." In the present chapter it is used throughout and is strictly correct when the axes referred to are orthogonal. If the axes are oblique, then much of what is said really refers to the items in a structure, not in a pattern : but the word " loading " is still used to avoid circumlocutions, and because the structure of the reference vectors is, except for a diagonal matrix multiplier, identical with the pattern of the factors.

multiplication of both the simple structure tables—both that from correlations and that from covariances—by the same orthogonal rotating matrix will leave their equivalence with regard to units unharmed. Simple structure is only one of an infinite number of positions which possess this property. But it is an easily identifiable one.

It is difficult to keep one's mind clear as to the meaning of this. Let me recapitulate. There are some processes of analysis which, while they give a perfect analysis in the sense of one which reproduces the correlations (or the co-variances) exactly, do not give the same analysis for the correlations as for the covariances. The factors they arrive at depend upon the units of measurement employed in the tests. Such, for example, are the principal components process and the centroid process. Such processes cannot be relied on to give straight away and without rotation, factors which can be called objective and scientific. Some processes, on the other hand, do give analyses which are independent of the units. One such is Lawley's, based on maximum likelihood. Another is Thurstone's simple-structure process, which, though it begins by using a centroid analysis, follows this by rotation of a certain kind.

But the principle of independence of units does not distinguish between these processes, which both satisfy it. Still less does it distinguish between systems of factors. For any one of the infinite number of such systems which can be got from either simple structure or Lawley's factors by rotation equally satisfies the principle. Indeed, there can really be no talk of a system of factors satisfying the principle. Any table of loadings whatever, obtained from correlations, has, of course, corresponding to it a system differing only in that the rows are multiplied by coefficients, a system which would correspond with covariances. The fact that no one has discovered a process which gives both is irrelevant. The argument is rather as follows. If a worker believes that he has found a process which gives the true psychological factors, then that process must be independent of the metric, and simple structure and maximum likelihood are both thus independent, though they do not, alas, agree. Nor must it be forgotten that analyses from correlations are in no way superior to those from covariances. Indeed, correlations *are* covariances, dependent upon as arbitrary a choice of units—namely standard deviations—as any other. But centroid axes in themselves, or principal components, without rotation, are clearly inadmissible, for they change with the units used. The chance that such axes are the true ones is infinitesimal, being dependent on the chance composition of the battery, and the system of units which chances to be used. Independence of metric is not sufficient to validate a process but it is necessary. Its absence does not prove a system of factors to be wrong, but it makes it certain that the process by which they have been arrived at does not in general give the true factors.

3. *Specifics.*—These form a fundamental problem in factorial analysis and yet they are practically never heard of in discussions of an analysis. It is reasonable enough to think that a test may require some trick of the intellect peculiar to itself, yet it is not obvious that these specific factors must be made as large and important as possible; and that is what the plan of minimizing the rank of a matrix does. The excess of factors over tests which inevitably, of course, results from postulating a specific in every test, means that the factors cannot be estimated with any great accuracy. Usually the accuracy is very low indeed. The determinate and the indeterminate parts of each of Thurstone's factors in *Primary Mental Abilities* can be found by post-multiplying Table 7 on his page 98 by Table 3 on his page 96. We find:

Factor	*Variance of the Estimated Part*	*Variance of the Indeterminate Part*
S . . .	·611	·389
P . . .	·616	·384
N . . .	·825	·175
V . . .	·662	·338
M . . .	·431	·569
W . . .	·439	·561
I . . .	·397	·603
R . . .	·600	·400
D . . .	·519	·481

The average for the nine factors is only $56\frac{1}{2}$ per cent. of the variance estimated. In other words the factor estimates have large probable errors in some cases as large as the estimates themselves. This has serious consequences, not to be overcome by more reliable tests.

Using unity for every diagonal element in the matrix of such a battery will give factors (supposing the same number of them to be taken out) which will not imitate the correlations quite so well, but which can be estimated accurately.

In fact, whether Hotelling's process or the centroid process is used, with unit communalities, each factor can be calculated exactly for a man, given his scores. By exactly we mean that they are as accurate as his scores are. Of course, in any psychological experiment the scores may not be accurate in the sense that they can be exactly reproduced by a repetition of the experiment. Apart from sheer blunders and clerical errors, there is the fact that a man's performance fluctuates from day to day. But these errors are common to any process of calculation which may be used on the scores. These are not the errors for which we are criticizing estimates of a man's factors. The point we are making is that factors based on communalities less than unity have a further, and large, *error of estimation*, whereas factors based on unit communalities (even if only one or two or a few are taken out) have no such further error of estimation.

If a few such factors taken out with unit communalities are then rotated (keeping them in the same space, i.e not changing their number) they still remain susceptible of exact estimation in a man.

As soon, however, as any fractions, minimum or not, are placed in the diagonal cells, we have thereby decided to use, in describing our tests, more orthogonal axes than there are tests ; for each test has then a specific factor, and there are in addition the common factors. This means in terms of our spatial model that none of the axes, neither the common factors nor the specific factors, are in the test space at all (except at the origin where they all cross). It is only about the test space, of dimensions equal to the number of tests, that we have any information from our battery. These axes are away in outer darkness and we cannot know them, but only their projections or shadows on the test space. Psychologists invariably confine their attention, after making an analysis using communalities, to the "common factor space," of a comparatively small number of dimensions, without, I think, being usually aware that this space is not in the test space at all. (Thurstone's "secondary factors," in their turn, are not even in the common factor space, for he uses what I might call secondary communalities.) The effect of all this is that the factors arrived at by an analysis which has begun by placing fractions in the diagonal cells can never be measured in any man, but only vaguely estimated, and with maximum vagueness if minimum communalities are used.

In itself the fact that factors can only be estimated and not accurately measured is, of course, not fatal. Throughout statistical work runs the idea of estimation in a realm outside that which is experimentally known, in a realm of more dimensions than that in which our measurements have been made. It is to allow for that that the device of "degrees of freedom" is used in the analysis of variance. But in factorial analysis the vagueness due to estimation is deliberately maximized, for reducing the rank of a matrix of correlations involves the simultaneous maximizing of the specific variances. In Section 3 of the previous chapter a brief reference was made to this fact that methods of factorizing which use communalities maximize the variance of the specific factors, by reason of minimizing the number of common factors. First take the case of the analysis of a hierarchical battery. As was illustrated in Chapter XX the analysis of such a battery into one general factor only, and specifics, gives the maximum variance possible to the specifics. The combined communalities of the tests are less in the two-factor analysis than in any other analysis. The mathematical expression of this is that the trace of the reduced correlation matrix, i.e. the sum of the cells of the principal diagonal, is a minimum.

It is true that certain exceptions to this statement are mathematically possible, but their occurrence in actual psychological work is a practical impossibility. They have been investigated by Ledermann (Ledermann, 1940), who finds, in the case of the hierarchical matrix, that an exception is only possible when one of the g saturations is greater than the sum of all the others. When the battery is of any size, this is most unlikely to occur : and almost always, when it did occur, the large saturation of one test would turn out to be greater than unity, which is not permissible (the Heywood case).

The same statement as the above, that the specifics are maximized, is also true in general. The communalities which give the matrix its lowest rank are in sum less than any other diagonal elements permissible. If smaller numbers are placed in the diagonal cells, the analysis fails unless factors with a loading of $\sqrt{-1}$ are employed, and such factors are, of course, inadmissible.

Here again there are possibly cases where the lowest rank is not accompanied by the lowest trace (i.e. the lowest sum of the communalities). But here again it seems certain that if such cases do exist, they are mathematical curiosities which would never occur in practice.

If specific factors of such large size have any psychological existence, what can they be ? Possibilities which will occur to us are first, that they are error factors—but errors or variations in the subject's performance are not likely to be entirely unique to one test. Secondly, they have been attributed to sampling errors in the coefficients of correlation—but these sampling errors are themselves correlated, and so give rise to false common factors, not to specific factors. Thirdly, they may be real mental factors, unique to that test, needed only by it. But what remarkable consequences follow if we accept that. I devise a brand-new test and, lo, in the mind of man there exists a specific ability to do that test and, moreover, an ability which is useless in every other activity. Further, every individual I meet possesses this specific ability in large or small amount. The idea in this form is really fantastic.

It would seem then that the specifics cannot be really unique, but only unique in this battery. This leads to the presumption that the tests of a battery possess specific factors only because there does not happen to be in the battery any other test to share the specific, or at least part of it, and prove it to be really one or more common factors. On this view, specifics will disappear when a test has been tried in a large number of batteries, or in a sufficiently large battery. Not only does this seem unlikely when one considers that in every battery the minimum communalities and maximum specifics are insisted on, but it also has peculiar consequences in regard to the number of primary factors. Consider a battery consisting of, say, two dozen tests, analysed into, say, seven common factors plus, of course, two dozen specifics. The latter, it must be remembered, are all orthogonal, all uncorrelated with one another. On the hypothesis that they are really factors which just do not happen to have found a partner, like wallflowers at a ball, there must exist at least two dozen other primary factors waiting to be discovered in a larger battery. And so with every battery of tests. The number of primary factors must be larger than all the tests hitherto invented, which does not seem to be parsimonious. I cannot help fearing that there is something wrong with the idea of reducing the matrix of correlation coefficients or co-variances to its lowest possible rank, and then calling the descriptive variates to which this leads "factors of the mind": something wrong with the whole idea of attributing as much as possible of the variance of a test to a unique factor, something wrong with the "parsimony" argument upon which all this is based. It leads to too many difficulties to which it is possible, but not, I think advisable, to shut one's eyes. Moreover, the reciprocity principle, which identifies factors and loadings obtained from correlating tests with loadings and factors obtained from correlating persons, works only when there are no specifics involved. I would like to see a number of existing squares of correlation coefficients re-analysed with full variance in each diagonal cell and the results considered. There would be no guessing of the communalities, and no repetitions or iterations of the calculation to determine them. Tests of significance of residues would be more easily made, and although rather more factors would be necessary before the residues became insignificant, they would have the advantage of more accurate estimation in any man. True, such factors would be confined to the particular test space of that battery, and admittedly a factor of the mind is not likely to be an exact composite of the tests of any one battery. But the point is an academic one, for the common-factor space in which communality factors exist, is just as much a creation of the particular battery as are axes determined within the battery space.

I must not be misunderstood as saying that no specific factors exist at all. What I am sceptical about is the procedure of making the specific factors in every battery as large as possible, by the automatic application of a mathematical device. That every test may well have some unique quality for any individual person seems conceivable, though I do not think this special feature of the test will be felt as a peculiarity by every person who tries the test. I think any such unique quality would be a blemish in the test, just as unreliability is a blemish, and that the psychologist should endeavour to make tests which are neither unreliable nor burdened with unique peculiarities. Probably he cannot avoid a certain amount of uniqueness, just as he cannot avoid a certain amount of unreliability. But I do not see the need for ascribing *maximum* uniqueness in order to reduce the number of common factors.

A critic may point out that, if even small truly unique parts of the tests are admittedly present, there will always be the need for the large number of specifics. Possibly so —but specifics of no great importance, if the tests are good ones ; specifics with an influence as unimportant as the causes are of the residuals which we in any case ignore after statistical testing.

It is true that by the use of communalities the total number of *loadings* to be estimated is reduced to a minimum. That way of putting the parsimony argument would be perhaps defensible. What I doubt is whether too high a price is not paid, since this same procedure maximizes the specifics, and decides their importance without any psychological consideration whatever being given to the question.

The practical conclusions I would draw from these considerations about the nature of specific factors are that a battery used for factorial analysis should be composed of tests of high communality in that battery: or that, if tests are admitted which by the mathematical principle of rank reduction are allotted low communalities, the psychologist should agree that these tests do draw, each of them, upon factors of the mind not represented elsewhere in the battery.

Such is the argument against minimum communalities. For them is the hope that some day, despite their drawbacks, the factors they lead to may prove to be something real, perhaps have some physiological basis. And their defender may plead that the estimates of these factors are as good as the estimates we find useful, in predicting educational or occupational efficiency.

4. *Oblique factors.*—I think it is pretty certain that Thurstone took to oblique factors because he wants simple structure at all costs. Certainly oblique factors make it much easier to reach simple structure—too easy, Reyburn and Taylor say. It will be found far more often than it really exists, they add. On the other hand, Thurstone can point to his box example and his trapezium example and say with truth that simple structure enabled him to find "realities," can say that the oblique simple structure is something more real, in the ordinary common-sense everyday use of the word, than the orthogonal second-order factors which are an alternative.

Other workers, not at all wedded to the ideas of simple structure, have also declared their belief in oblique factors, e.g. Raymond Cattell, and, I think, many who feel inclined to work in terms of "clusters." In ordinary life, weight and height are both measures of something real, although they are correlated. We could analyse them into two uncorrelated factors *a* and *b*, or into three for that matter, but certainly no one would use these in ordinary life. It is, however, just conceivable that some pair of hormones (say) might be found which corresponded, not one of them to height and one to weight, but one to orthogonal factor *a* and another to orthogonal factor *b*. It is far too early to state anything more than a preference for orthogonal or oblique factors. Opinion is turning, I think, toward the acceptance of the latter.

MATHEMATICAL APPENDIX

MATHEMATICAL APPENDIX

Paragraphs

1. *Textbooks on matrix algebra.*—Some knowledge of matrix algebra is assumed, such as can be gained from the mathematical introduction to L. L. Thurstone's *Multiple Factor Analysis* (Chicago, 1947); Turnbull and Aitken's *Theory of Canonical Matrices*, Chapter I (London and Glasgow, 1932); H. W. Turnbull's *The Theory of Determinants, Matrices, and Invariants*, Chapters I–V (London and Glasgow, 1929); and M. Bôcher's *Introduction to Higher Algebra*, Chapters II, V, and VI (New York, 1936).

I have adopted Thurstone's notation in Sections 19 and 19*a* of the mathematical appendix, and in Chapters XI, XII, and XIII in describing his work. But I have not made the change elsewhere because readers would then be incommoded in consulting my own former papers.

The chief differences are as follows:

My M is Thurstone's F, for centroid factors, my Z is Thurstone's $S \div \sqrt{N}$, and my F is Thurstone's $P \div \sqrt{N}$.

2. *Matrix notation.*—Let X be the matrix of raw scores of p persons in n tests, with n rows and p columns; and when normalized by rows, let it be denoted by Z. The letters z and Z in the *text* of this book mean *standardized* scores, which are used in practical work, but in this appendix they mean *normalized* scores, so that—

$$ZZ' = R \quad . \quad . \quad . \quad (1)$$

the matrix of correlations between n tests.

For many purposes it is convenient to think of solid matrices like Z as column (or row) vectors of which each element represents a row (or column). Thus Z can be thought of as a column vector z, of which each element represents in a collapsed form a row of test scores. Thus with three tests and four persons—

$$z = \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} = \begin{bmatrix} z_{11} & z_{12} & z_{13} & z_{14} \\ z_{21} & z_{22} & z_{23} & z_{24} \\ z_{31} & z_{32} & z_{33} & z_{34} \end{bmatrix} = Z \quad . \quad (2)$$

In the theory of mental factors each score is represented as a loaded sum of the normalized factors f, the loadings being different for each test, i.e.—

$$z = Mf \text{ (specification equations)} \quad . \quad (3)$$

where M is the matrix of loadings and f the vector of v factors, collapsed into a column from F, the full matrix, of dimensions $v \times p$.

We note that p = number of persons,
n = number of tests,
v = number of factors.

The dimensions of M are $n \times v$. Equation (3) represents n simultaneous equations, and the form $Z = MF$ represents np simultaneous equations.

We now have—

$$R = ZZ' = (MF)(MF)' = MFF'M' \quad . \quad (4)$$

If the factors are orthogonal, we have—

$$FF' = I \quad . \quad . \quad . \quad . \quad . \quad (5)$$

the unit matrix, and therefore—

$$R = MM' \quad . \quad . \quad . \quad . \quad . \quad (6)$$

The resemblance in shape between this and—

$$R = ZZ' \quad . \quad . \quad . \quad . \quad . \quad (1)$$

leads to a parallelism between formulæ concerning persons and factors (Thomson, 1935*b*, 75 ; Mackie, 1928*a*, 74, and 1929, 34).

3. *Spearman's Theory of Two Factors* assumes that M is of the special form—

$$M = \begin{bmatrix} l_1 & m_1 & . & . & . \\ l_2 & . & m_2 & . & . \\ . & . & . & . & . \\ l_n & . & . & . & m_n \end{bmatrix}, \ l^2 + m^2 = 1 \quad . \quad (7)$$

and therefore—

$$R = ll' + M_1^2 \quad . \quad . \quad . \quad (8)$$

where M_1 is the diagonal matrix which forms the right-hand end of M, and l is the first column of M. In this form it is clear that R is of rank 1 except for its principal diagonal. Its component ll' is the " reduced correlational matrix " of the Spearman case, and is entirely of rank 1. The elements $l_1^2, l_2^2, \ldots l_n^2$, which form the principal diagonal of ll', are called " communalities."

4. *Multiple common factors.*—When more than one common factor is present, M takes the form—

$$M = (M_0 \mid M_1) \quad . \quad . \quad . \quad (9)$$

where M_0 is the matrix of loadings of the common factors, represented in the Spearman case by the simple column l. We have then—

$$R = MM' = M_0M_0' + M_1^2 \quad . \quad . \quad (10)$$

where the " reduced correlation matrix " M_0M_0' is of rank r, the number of *common* factors, and is identical with R except for having " communalities " in its principal diagonal.

5. *Orthogonal rotations.*—If we express the v factors f in terms of w new factors φ by the equation—

$$f = A\varphi \quad . \quad . \quad . \quad . \quad . \quad (11)$$

where A is a matrix of v rows and w columns, we have—

$$z = Mf = MA\varphi \quad . \quad . \quad . \quad (12)$$

an expression of the tests z as linear loaded sums of a different set of factors, with a matrix of loadings MA.

If—

$$AA' = I \quad . \quad . \quad . \quad . \quad . \quad (13)$$

the new factors φ are orthogonal like the old ones. They can be as numerous as we like, but not less than the number of tests unless the matrix R is singular. (12) represents a rigid rotation of the orthogonal axes f into new positions, with dimensions added or abolished.

6. *The sampling theory.*—The following transformation is of interest as showing the connexion between the Theory of Two Factors and the Sampling Theory (Thomson, 1935*b*, 85). We shall write it out for three tests only, but it is quite general. Consider the orthogonal matrix:

$$\left[\begin{array}{c|ccc|ccc|c}
lll & mll & lml & llm & mml & mlm & lmm & mmm \\
\hline
mll & -lll & mml & mlm & -lml & -llm & mmm & -lmm \\
lml & mml & -lll & lmm & -mll & mmm & -llm & -mlm \\
llm & mlm & lmm & -lll & mmm & -mll & -lml & -mml \\
\hline
mml & -lml & -mll & mmm & lll & -lmm & -mlm & llm \\
mlm & -llm & mmm & -mll & -lmm & lll & -mml & lml \\
lmm & mmm & -llm & -lml & -mlm & -mml & lll & mll \\
\hline
mmm & -lmm & -mlm & -mml & llm & lml & mll & -lll
\end{array}\right] \quad (14)$$

wherein the omitted subscripts 1, 2, and 3 are to be understood as existing always in that order, so that mll means $m_1l_2l_3$.

If we take for A in Equation (12) the first four rows of this orthogonal matrix, and for M the Spearman form (7) with three tests, the result is to transfer to eight new factors, yielding:

$$\begin{aligned}
z_1 &= l_2l_3\varphi_1 + m_2l_3\varphi_3 + l_2m_3\varphi_4 + m_2m_3\varphi_7 \\
z_2 &= l_1l_3\varphi_1 + m_1l_3\varphi_2 + l_1m_3\varphi_4 + m_1m_3\varphi_6 \\
z_3 &= l_1l_2\varphi_1 + m_1l_2\varphi_2 + l_1m_2\varphi_3 + m_1m_2\varphi_5
\end{aligned} \quad . \quad . \quad (15)$$

Each z is here in normalized units. If, however, we change to new units by multiplying the three equations by l_1, l_2, and l_3 respectively, we have:

$$\begin{aligned} l_1 z_1 &= l_1 l_2 l_3 \varphi_1 + l_1 m_2 l_3 \varphi_3 + l_1 l_2 m_3 \varphi_4 + l_1 m_2 m_3 \varphi_7 \\ l_2 z_2 &= l_1 l_2 l_3 \varphi_1 + m_1 l_2 l_3 \varphi_2 + l_1 l_2 m_3 \varphi_4 + m_1 l_2 m_3 \varphi_6 \\ l_3 z_3 &= l_1 l_2 l_3 \varphi_1 + m_1 l_2 l_3 \varphi_2 + l_1 m_2 l_3 \varphi_3 + m_1 m_2 l_3 \varphi_5 \end{aligned} \quad . \quad . \quad (16)$$

and the variates $l_1 z_1$, $l_2 z_2$, and $l_3 z_3$ are now susceptible of the explanation that each is composed of $l_i^2 N$ small equal components drawn at random from a pool of N such components, all-or-none in nature. In that case $l_1^2 l_2^2 l_3^2 N$ components would probably appear in all three drawings (φ_1); $l_1^2 l_2^2 m_3^2 N$ components would probably appear in the first two drawings, but not in the third (φ_4); and so on down to $m_1^2 m_2^2 m_3^2$ components, which would not appear at all (φ_8, which is missing from the equations).

The transformation can, of course, be reversed, and the sampling theory equations converted into the two-factor equations.

7. *Hotelling's "principal components"* are the principal axes of the ellipsoids of equal density—

$$z' R^{-1} z = \text{constant} \quad . \quad . \quad . \quad (17)$$

when the *test* vectors are orthogonal axes (Hotelling, 1933). To find the principal axes involves finding the latent roots of R^{-1}. The Hotelling process consists of (*a*) a rotation of the axes from the orthogonal text axes to the directions of the principal axes; and (*b*) a set of strains and stresses along these new axes to standardize the factors, making the ellipsoid spherical and the original axes oblique. The transformation from the tests to the Hotelling factors γ being from Equation (3)—

$$z = M\gamma \quad (M \text{ square})$$

the ellipsoids (17) become—

$$\text{constant} = z' R^{-1} z = \gamma'(M' R^{-1} M)\gamma = \gamma'\gamma \quad . \quad (18)$$

since they become spheres. Therefore we must have—

$$M' R^{-1} M = I \quad . \quad . \quad . \quad (19)$$

The locus of the mid points of chords of $z' R^{-1} z$ whose direction cosines are h' is the plane $h' R^{-1} z = 0$, and if this is a principal plane it is at right angles to the chords it bisects, i.e.—

$$h' R^{-1} = \lambda h'$$

which has non-trivial solutions only for—

$$| R^{-1} - \lambda I | = 0$$

the roots λ of which are the "latent roots" of R^{-1}, while each h' is a "latent vector."

Now, if H is the matrix of normalized latent vectors of R^{-1}, we have—

$$H'R^{-1}H = \Lambda$$

where Λ is the diagonal matrix of the latent roots of R^{-1}; so that a solution for M corresponding to rotation to the principal axes and subsequent change of units to give a sphere is seen to be—

$$M = H\Lambda^{-\frac{1}{2}} \quad . \quad . \quad (20)$$

The latent vectors of R are the same as those of R^{-1}, or of any power of R, and Hotelling's process described in the text (Chapter VII) finds the latent roots (forming the diagonal matrix D) and the latent vectors (forming H) of R. We then have—

$$M = HD^{\frac{1}{2}} \quad . \quad . \quad . \quad (21)$$

For the convergence of the process, see Hotelling's paper of 1933, pages 14 and 15.

Since in Hotelling analyses M is square, we can write—

$$\begin{aligned} \gamma &= M^{-1}z = (HD^{\frac{1}{2}})^{-1}z \\ &= D^{-\frac{1}{2}}H^{-1}z = D^{-1}(D^{\frac{1}{2}}H')z = D^{-1}M'z \quad . \quad (22) \end{aligned}$$

Each factor γ, that is, can be found from a *column* of the matrix M, divided by the corresponding latent root, used as loadings of the test scores z.

8. *The pooling square.*—If the matrix of correlations of $a + b$ variates is :

$$\begin{array}{c|c} R_{aa} & R_{ab} \\ \hline R_{ba} & R_{bb} \end{array} \quad . \quad . \quad . \quad (23)$$

and if the standardized variates a are multiplied by weights u, the standardized variates b by weights w, and each set of scores summed to make two composite scores, the resulting variances and covariances are :

$$\begin{array}{c|c} u'R_{aa}u & u'R_{ab}w \\ \hline w'R_{ba}u & w'R_{bb}w \end{array} \quad . \quad . \quad . \quad (24)$$

as can be seen by writing out the latter expressions at length. The battery intercorrelation is therefore—

$$\frac{u'R_{ab}w \quad \text{or} \quad w'R_{ba}u}{\sqrt{(u'R_{aa}u \times w'R_{bb}w)}} \quad . \quad . \quad (25)$$

If weights are applied to *raw* scores, each applied weight must be multiplied by each pre-existing standard deviation, in (25).

If there is only one variate in the *a* team, (25) becomes—

$$\frac{w'r_{ba}}{\sqrt{(w'R_{bb}w)}} \quad . \quad . \quad . \quad (26)$$

where r_{ba} represents a whole column of correlation coefficients. The values of w for which this reaches its maximum value will satisfy the equation—

$$\frac{\delta}{\delta w} \frac{w'r_{ba}}{\sqrt{(w'R_{bb}w)}} = 0 \quad . \quad . \quad (27)$$

that is—

$$w = \text{a scalar} \times R_{bb}^{-1}r_{ba} \quad . \quad . \quad (28)$$

consistent with the ordinary method of deducing regression coefficients.

9. *The regression equation.*—If z_0 is the one variate in the *a* team, and z are the *b* team, and if —

$$\hat{z}_0 = w'z \quad . \quad . \quad . \quad . \quad (29)$$

we wish to make $S(z_0 - \hat{z}_0)^2$ a minimum, that is—

$$\frac{\delta}{\delta w} S(z_0 - w'z)^2 = 0$$

$$Sz_0z' = w'Szz'$$

$$w' = r_{ab}'R_{bb}^{-1}$$

$$\hat{z}_0 = r_{ab}'R_{bb}^{-1}z \quad . \quad . \quad . \quad (30)$$

If **R** is the matrix of correlations of all the tests including z_0, the regression estimate of any one of the tests from a weighted sum of the others is given by—

$$\text{determinant } \mathbf{R}_z = 0 \quad . \quad . \quad (31)$$

where $\mathbf{R}_z$ is **R** with the row corresponding to the variate to be estimated replaced by the row of variates.

9*a*. *Relations between two sets of variates.*—(Hotelling 1935*a*, 1936, M. S. Bartlett 1948). If two sets of variates have correlation coefficients—

$$\begin{array}{c|c} R_{aa} & R_{ab} \\ \hline R_{ba} & R_{bb} \end{array} \quad \text{or} \quad \begin{array}{c|c} A & C \\ \hline C' & B \end{array}$$

and if the variates of the B team are fitted with weights b, then the correlations of the B team, thus weighted, with the separate tests of the A team are given by—

$$r = \frac{C'b}{\sqrt{b'Bb}} \qquad \qquad (31.1)$$

and the square of the correlation coefficient between the two teams is then—

$$\frac{b'CA^{-1}C'b}{b'Bb} = \lambda \qquad \qquad (31.2)$$

The maximum intercorrelation, and other points of inflexion in λ, will be given by—

$$d\lambda/db = 0$$

$$\text{i.e. } (CA^{-1}C' - \lambda B)b = 0 \qquad \qquad (31.3)$$

a set of homogeneous equations in b. We must therefore have—

$$\left| CA^{-1}C' - \lambda B \right| = 0 \qquad \qquad (31.4)$$

an equation for λ with as many non-zero roots as the number of variates in the smaller team. For any one of these roots λ, the weights b are proportional to the co-factors of any row of $(CA^{-1}C' - \lambda B)$. The corresponding weights a for the A team are then found by condensing the team B (using weights b) to a single variate and carrying out an ordinary regression calculation.

The result is to " factorize " each team into as many orthogonal axes as there are variates. These axes are related to one another in pairs corresponding to the roots λ. Each axis is orthogonal to all the others except its own opposite number in the space of the other team, arising from the same root λ as it does, to which axis it is inclined at an angle arccos $\sqrt{\lambda}$. Where one team has m more variates than the other, m of the roots will be zeros and the corresponding axes will be at right angles to the whole space of the other team. This form of factorizing has been called by M. S. Bartlett (1948) *external* factorizing, since the position of the " factors " or orthogonal axes in each team, in each space, is dictated by the other team.

The weightings corresponding to the largest root give the closest possible correlation of the two weighted teams. If the two teams are duplicate forms of the same tests, this is the maximum attainable battery or team reliability (Thomson 1940, 1947, 1948). In this case Peel (*Nature*, 1947) has shown that a simpler equation than 31·4 gives the required roots. If $\lambda = \mu^2$ Peel's equation is—

$$|C - \mu A| = 0 \quad . \quad . \quad . \quad (31.5)$$

where A differs from C only in the diagonal elements, which in A are unities but in C are reliabilities r_{ii} of the individual tests.

Green (1950) gives a transformation of this equation which enables Hotelling's iterative process (see Chapter VII) to be used to find μ, the maximum battery reliability. For the diagonal elements $r_{ii} - \mu$ of the matrix $(C - \mu A)$, Green writes—

$$\left[\frac{r_{ii}}{1 - r_{ii}} - \frac{\mu}{1 - \mu}\right](1 - r_{ii})(1 - \mu)$$

when 31·5 becomes equivalent to—

$$|DCD - \beta I| = 0 \quad . \quad . \quad . \quad (31.6)$$

wherein D is a diagonal matrix with elements $(1 - r_{ii})^{-\frac{1}{2}}$, I is the unit matrix, and $\beta = \mu/(1-\mu)$. The latent vector V corresponding to the largest latent root of DCD can then be found by Hotelling's process, and the best weights for maximum battery reliability are proportional to $DV = W$. The maximum reliability thus attained is—

$$\mu = W'CW/W'AW$$

10. *Regression estimates of factors.*—When in the specifications—

$$z = Mf \quad . \quad . \quad . \quad . \quad (3)$$

the factors outnumber the tests, they cannot be measured but only estimated. To all men with the same set of scores z will be attributed the same set of estimated factors f, though their " true " factors may be different. The regression method of estimation minimizes the squares of the discrepancies between $\hat{f}$ and f, summed over the men. The regression equation (31) will be for one factor f_i—

$$\begin{vmatrix} \hat{f}_i & z' \\ m_i & R \end{vmatrix} = 0 \quad . \quad . \quad . \quad (32)$$

where m_i is a column of M. Expanding, we have—

$$\hat{f}_i = m_i' R^{-1} z$$

and in general—

$$\hat{f} = M' R^{-1} z \quad . \quad . \quad . \quad (33)$$

or, separating the common factors and the specifics—

$$\hat{f}_0 = M_0' R^{-1} z \quad . \quad . \quad . \quad (34)$$

$$\hat{f}_1 = M_1 R^{-1} z \quad . \quad . \quad . \quad (35)$$

the latter of which shows that we know the *proportionate* weights for each specific (the rows of R^{-1}) even before we know whether that specific exists (Wilson, 1934, 194). The matrix of covariances of the estimated factors is—

$$M' R^{-1} M = \begin{bmatrix} M_0' R^{-1} M_0 & M_0' R^{-1} M_1 \\ M_1 R^{-1} M_0 & M_1 R^{-1} M_1 \end{bmatrix} \quad . \quad . \quad (36)$$

a square idempotent matrix of order equal to the number of factors, but trace only equal to the number of tests.

For one common factor, (34) reduces to Spearman's estimate—

$$\hat{g} = \frac{1}{1+S} \Sigma \frac{r_{ig} z_i}{1 - r_{ig}^2} \quad . \quad . \quad (34a)$$

where $$S = \Sigma \frac{r_{ig}^2}{1 - r_{ig}^2}$$

while $K = M_0' R^{-1} M_0$ in (36) reduces to $S/(1 + S)$, the variance of $\hat{g}$.

10*a*. *Ledermann's short cut* (1938*a*, 1939*b*).—The above requires the calculation of the reciprocal of the large square matrix R. Ledermann's short cut only requires the reciprocal of a matrix of order equal to the number of common factors. As long as the factors are orthogonal we have—

$$R = M_0 M_0' + M_1^2 \quad . \quad . \quad . \quad (10)$$

and the identity

$$M_0' M_1^{-2} (M_0 M_0' + M_1^2) = (M_0' M_1^{-2} M_0 + I) M_0'$$
$$= (J + I) M_0' \text{ say.}$$

Premultiplying by $(I + J)^{-1}$ and postmultiplying by R^{-1} we reach $(I + J)^{-1}M_0'M_1^{-2} = M_0'R^{-1}$ . . . (36.1) and the left-hand quantity can then be used in Equation (34).

This short cut requires modification when the factors are oblique. See Equations (70.1) to (70.4) below.

11. *Direct and indirect vocational advice.*—If z_0 is an occupation and z a battery of tests, the estimate of a candidate's occupational ability is—

$$\hat{z}_0 = r_0'R^{-1}z \quad . \quad . \quad . \quad (37)$$

where the r_0 are the correlations of the occupation with the tests. If z_0 can be specified in terms of the common factors of z, and a specific s_0 independent of z, then an indirect estimate of z_0 *via* the estimated f_0 is possible. We have—

$$z_0 = m_0'f_0 + s_0 \quad . \quad . \quad . \quad (38)$$

where m_0' is a row of occupation loadings for the common factors f_0 of z, and also—

$$\hat{f}_0 = M_0'R^{-1}z$$

Substitution in (38), assuming an average s_0 $(= 0)$ gives—

$$\hat{z}_0 = m_0'M_0'R^{-1}z \quad . \quad . \quad . \quad (39)$$

But—

$$m_0'M_0' = r_0' \quad . \quad . \quad . \quad (40)$$

and (39) is identical with (37) (Thomson, 1936*a*). If, however, s_0 is not independent of the specifics s of the battery, (40) will not hold, and the estimate (39) made *via* an estimation of the factors will not agree with the correct estimate (37).

12. *Computation methods.*—The "Doolittle" method of computing regression coefficients is widely used in America (Holzinger, 1937*a*, 32). Aitken's method, used and explained in the text, is in the present author's opinion superior (Aitken, 1937*a* and *b*, with earlier references). Regression calculations and many others are all special cases of the evaluation of a triple matrix product $XY^{-1}Z$, where Y is square and non-singular, and X and Z may

be rectangular. The Aitken method writes these matrices down in the form—

$$\begin{array}{c|c} Y & -Z \\ \hline X & . \end{array}$$

and applies pivotal condensation until all entries to the left of the vertical line are cleared off. All pivots must originate from elements of Y. By giving X and Z special values (including the unit matrix I) the most varied operations can be brought under the one scheme.

13. *Bartlett's estimate of factors.*—We have $z = M_0 f_0 + M_1 f_1$, where f_0 and f_1 are column vectors of the common and specific factors respectively and M_1 is a diagonal matrix. Bartlett now makes the estimates $\check{f}_0$ such as will minimize the sum of the squares of each person's specifics over the battery of tests, i.e.—

$$\frac{\delta}{\delta f_0}(f_1' f_1) = 0$$

or—

$$\left(\frac{\delta f_1}{\delta f_0}\right)' f_1 = 0$$

i.e.—

$$\begin{aligned} (-M_1^{-1}M_0)'(M_1^{-1}z - M_1^{-1}M_0 f_0) &= 0 \\ M_0'M_1^{-2}z &= M_0'M_1^{-2}M_0 f_0 \\ &= J f_0, \text{ say} \\ \check{f}_0 &= J^{-1}M_0'M_1^{-2}z \ . \quad (41) \end{aligned}$$

(Bartlett, 1937*a*, 100.)

One could also find the estimated specifics as—

$$\check{f}_1 = (I - M_1^{-1}M_0 J^{-1} M_0' M_1^{-1}) M_1^{-1} z \quad . \quad . \quad (42)$$

Substituting—

$$z = [M_0 \mid M_1] \begin{bmatrix} f_0 \\ \hline f_1 \end{bmatrix}$$

we get for the relation between $\check{f}$ and f—

$$\begin{bmatrix} \check{f}_0 \\ \hline \check{f}_1 \end{bmatrix} = \left[\begin{array}{c|c} I & J^{-1}M_0'M_1^{-1} \\ \hline . & I - M_1^{-1}M_0J^{-1}M_0'M_1^{-1} \end{array}\right] \begin{bmatrix} f_0 \\ \hline f_1 \end{bmatrix} = Af \quad (43)$$

and for the covariances of $\check{f}$ we get—

$$AA' = \left[\begin{array}{c|c} I + J & . \\ \hline . & I - M_1^{-1}M_0J^{-1}M_0'M_1^{-1} \end{array}\right] \quad . \quad (44)$$

The error variances and covariances of the common factors are—

$$(\check{f}_0 - f_0)(\check{f}_0 - f_0)' = J^{-1}M_0'M_1^{-1}(f_1f_1')M_1^{-1}M_0J^{-1}$$
$$= J^{-1}M_0'M_1^{-2}M_0J^{-1} = J^{-1} \quad . \quad (45)$$

(Bartlett, 1937*a*, 100.)

When there is only one common factor, J becomes the familiar quantity—

$$J = S = \Sigma \frac{r_{ig}^2}{1 - r_{ig}^2}$$

(Bartlett, 1935, 200.)

As was first noted by Ledermann *—

$$I + J^{-1} \equiv (M_0'R^{-1}M_0)^{-1} = K^{-1} \quad . \quad (46)$$

(quoted by Thomson, 1938*a*); and using this we see that the back estimates of the original scores from the *regression* estimates $\hat{f}_0$ are identical with the insertion of Bartlett's estimates $\check{f}_0$ in the common-factor part of the specification equations, viz.—

$$M_0K^{-1}M_0'R^{-1}z = M_0J^{-1}M_0'M_1^{-2}z \quad . \quad . \quad (47)$$

(Thomson, 1938*a*.)

Bartlett has pointed out that, using the same identity, in the form $K \equiv J(I - K)$, it is easy to establish the reversible relation between his estimates and regression estimates—

$$\hat{f}_0 = K\check{f}_0, \; \check{f}_0 = K^{-1}\hat{f}_0 \quad . \quad . \quad (48)$$

(Bartlett, 1938)

and he summarizes their different interpretation and properties by the formulæ—

$$E\{\hat{f}_0\} = E\{f_0\} = 0, \; E\{(\hat{f}_0 - f_0)(\hat{f}_0 - f_0)'\} = I - K \quad (49)$$

$$E_1\{\check{f}_0\} = f_0, \; E_1\{(\check{f}_0 - f_0)(\check{f}_0 - f_0)'\} = J^{-1}$$
$$= K^{-1}(I - K) \quad . \quad (50)$$

where E denotes averaging over all persons, E_1 over all possible sets of tests (comparable with the given set in regard to the amount of information on the group factors f_0).

14. *Indeterminacy*.—The fact that estimated factors, if the factors outnumber the tests, necessarily have less than

* Letter of October 23, 1937, to Thomson.

unit variance has sometimes been expressed in the case of one common factor by postulating an indeterminate vector i whose variance completes unity. This i may be regarded as the usual error of estimation, and is a function of the specific abilities (Thomson, 1934, *B.J.P.*, **25**, 92). That $M'R^{-1}M$ in Equation (36) is of rank less than its order also expresses the indeterminacy, and allows the factors to be rotated to different positions which nevertheless fulfil all the required conditions. In the hierarchical case the transformation which effects this is (Thomson, 1935*a*)—

$$f = \bar{B}\varphi \quad . \quad . \quad . \quad . \quad (51)$$

where $\bar{B}$ means the required number of rows of—

$$B = I - 2qq'/q'q \quad . \quad . \quad . \quad (52)$$

in which—

$$q_i = l_i/m_i \text{ (see Equation 7)} \quad . \quad (53)$$

as far as there exist tests, after which q is arbitrary.

For—

$$z = Mf = M\bar{B}\varphi = M\varphi$$

since—

$$M\bar{B} = M \quad . \quad . \quad . \quad (54)$$

and z is thus expressed by identical specification equations in terms of new factors φ. For such transformations in the case of multiple factors see Thomson, 1936*a*, 40; and Ledermann, 1938*c*.

If the matrix M is divided into the part M_0 due to common factors and the part M_1 due to specifics, as in Equation (9), then Ledermann shows that if U is any orthogonal matrix of order equal to the number of common factors, the matrix—

$$B = I - Q(Q'Q)^{-\frac{1}{2}}(I - U)(Q'Q)^{-\frac{1}{2}}Q'$$

wherein—
$$Q = \begin{bmatrix} -I \\ M_1{}^{-1}M_0 \end{bmatrix}$$

will satisfy the equation—

$$MB = M$$

Indeterminacy is entirely due to the excess of factors over tests, i.e. to the fact that the matrix of loadings M

is not square. It can be in theory abolished by adding a new test which contains no new factor, not even a new specific; or a set of new tests which add fewer factors than their number, so that M becomes square (Thomson, 1934*b*; 1935*a*, 253). In the case of a hierarchy each of these tests singly will conform to the hierarchy, so that their saturations l can be found; but jointly they break the hierarchy. If they add no new factors, g can then be found without any indeterminacy.

15. *Finding* g *saturations from an imperfectly hierarchical battery.*—The Spearman formula given in Chapter III, Section 5, is the most usual method. A discussion of other methods will be found in Burt (1936, 283–7). See also Thomson (1934*a*, 370), for an iterative process modified from Hotelling.

16. *Sampling errors of tetrad-differences.*—The formulæ (16) and (16A) given in the text are both approximations, but appear to be very good approximations. The primary papers are Spearman and Holzinger (1924 and 1925). Critical examination of the formulæ have been made by Pearson and Moul (1927), and Pearson, Jeffery, and Elderton (1929). Wishart (1928) has considered a quantity P which is equal to $P'N^2/(N-1)(N-2)$, where P' is the tetrad-difference of the covariances a instead of the correlations, and obtained an *exact* expression for the standard deviation σ of P—

$$(N-2)\sigma^2 = \frac{N+1}{N-1} D_{12}D_{34} - D + 3D_{13}D_{31} \quad . \qquad (55)$$

where the D's are determinants of the following matrix and its quadrants:

$$\left[\begin{array}{cc|cc} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ \hline a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{array}\right]$$

But approximate assumptions are necessary when the standard deviation of the ordinary tetrad-difference of the correlations is deduced from that of P. The result for the variance of the tetrad-difference is—

$$\frac{N+1}{(N-1)(N-2)}(1-r_{12}^2)(1-r_{34}^2)-R \quad . \quad (56)$$

where R is the 4×4 determinant of the correlations.

17. *Selection from a multivariate normal population.*—The primary papers are those of Karl Pearson (1902 and 1912). The matrix form given in the text (Chapter XIX, Section 2) is due to Aitken (1934), who employed Soper's device of the moment-generating function, and made a free use of the notation and methods of matrices. A variant of it which is sometimes useful has been given by Ledermann (Thomson and Ledermann, 1938) as follows. If the original matrix is subdivided in any symmetrical manner :

$$\begin{bmatrix} R_{pp} & R_{pq} & R_{ps} & R_{pt} & . \\ R_{qp} & R_{qq} & R_{qs} & R_{qt} & . \\ R_{sp} & R_{sq} & R_{ss} & R_{st} & . \\ R_{tp} & R_{tq} & R_{ts} & R_{tt} & . \\ . & . & . & . & . \end{bmatrix}$$

and R_{pp} is changed by selection to V_{pp}, then each resulting sub-matrix, including V_{pp} itself, is given by the formula—

$$\left.\begin{aligned} V_{\alpha\beta} &= R_{\alpha\beta} - R_{\alpha p}E_{pp}R_{p\beta} \\ \text{where—} \quad E_{pp} &= R_{pp}^{-1} - R_{pp}^{-1}V_{pp}R_{pp}^{-1} \end{aligned}\right\} \quad . \quad (57)$$

17*a*. *Maximum likelihood estimation.*—The maximum likelihood equations for estimating factor loadings (Lawley, 1940, 1941, 1943*b*) may be expressed fairly simply in the notation of previous sections. It is necessary, however, to distinguish between the matrix of observed correlations, which we shall denote by R_0, and the matrix—

$$R = M_0M_0' + M_1^2$$

which represents that part of R_0 which is " explained " by the factors.

The equations may then be written—

$$M_0' = M_0'R^{-1}R_0 \quad . \quad . \quad . \quad (58)$$

These are not very suitable for computational work. It may, however, be shown that—

$$M_0'R^{-1} = (I - K)M_0'M_1^{-2} = (I + J)^{-1}M_0'M_1^{-2} \quad (59)$$

where, as before,

$$K = M_0'R^{-1}M_0, \quad J = M_0'M_1^{-2}M_0$$

Hence our equations may be transformed into the form—

$$M_0' = (I + J)^{-1}M_0'M_1^{-2}R_0 \quad . \quad . \quad (60)$$

or alternatively,

$$M_0' = J^{-1}(M_0'M_1^{-2}R_0 - M_0') \quad . \quad (61)$$

When there are two or more general factors the above equations will have an infinite number of solutions corresponding to all the possible rotations of the factor axes. A unique solution may, however, be found such as to make J a diagonal matrix.

Finally, if we put—

$$L = M_0'M_1^{-2}R_0 - M_0'$$
$$V = LM_1^{-2}M_0,$$

then, from the last set of equations

$$V = JM_0'M_1^{-2}M_0 = J^2$$

Hence we have—

$$M_0' = V^{-\frac{1}{2}}L \quad . \quad . \quad . \quad . \quad (62)$$

These equations have been found the most convenient in practice, since they can be solved by an iterative process. When first approximations to M_0 and M_1 have been obtained, they can be used to provide second approximations by substitution in the right-hand side.

18. *Reciprocity of loadings and factors in persons and traits* (Burt, 1937*b*).—Let W be a matrix of scores centred both by rows and columns. Its dimensions are traits $\times$ persons $(t \cdot p)$, and its rank is r where r is smaller than both t and p in consequence of the double centring. The two matrices of covariances are WW' for traits and $W'W$ for persons, and by a theorem first enunciated by Sylvester in 1883 (independently discovered by Burt), their non-zero latent roots are the same. If their dimensions differ, i.e. $t \neq p$, the larger one will have additional zero roots.

Let the non-zero roots form the diagonal matrix D. Then the principal axes analyses are :

$$W = H_1 D^{\frac{1}{2}} F_1, \text{ dimensions } (t \,.\, r)(r \,.\, r)(r \,.\, p)$$

and $W' = H_2 D^{\frac{1}{2}} F_2$, dimensions $(p \,.\, r)(r \,.\, r)(r \,.\, t)$

where H_1 and H_2 are the latent vectors of WW' and $W'W$, while F_1 is the matrix of factors possessed by persons, F_2 that of factors possessed by traits. From the analysis of W we have, taking the transpose—

$$W' = F_1' D^{\frac{1}{2}} H_1', \text{ dimensions } (p \,.\, r)(r \,.\, r)(r \,.\, t)$$

and comparison of this with the former expression for W' makes the reciprocity of H_2 and F_1', F_2 and H_1', evident.

19. *Oblique factors. Structure and pattern.*—In Thurstone's notation, which we shall follow in this paragraph, the matrix M of our equation (3), when it refers to centroid factors, is called F. Our equation (3) becomes in his notation—

$$s = Fp$$

Since centroid factors are orthogonal, F is both a pattern and a structure. The structure is the matrix of correlations between tests and factors, i.e. :

Structure $= sp' = (Fp)p' = F(pp') = FI = F =$ Pattern.

When the factors are oblique, however, this is not the case. In that case, Structure = Pattern × matrix of correlations between the factors.

Thurstone turns the centroid factors to a new set of positions (still within the common-factor space, and in general oblique to one another) called reference vectors. The rotating matrix is Λ, and

$$V = F\Lambda \quad . \quad . \quad . \quad (63)$$

is the structure on the reference vectors. The cosines of the angles between the reference vectors are given by $\Lambda'\Lambda$. V is not a pattern. Its rows cannot be used as coefficients in equations specifying a man's scores in the tests, given his scores in the reference vectors. The pattern on the reference vectors would not have those zeros which are found in V.

The primary factors are the lines of intersection of the hyperplanes which are at right angles to the reference vectors, taken $(r-1)$ at a time where r is the number of common factors, the number of dimensions in the common-factor space. They are defined, therefore, by the equations of the hyperplanes, taken (r − 1) at a time. These equations are

$$\Lambda' x = O \quad . \quad . \quad . \quad (64)$$

where x is a column vector of co-ordinates along the centroid axes. The direction cosines of the intersections of these hyperplanes taken $(r-1)$ at a time are therefore proportional to the elements in the columns of $(\Lambda')^{-1}$, and to make them into direction cosines this has to have its columns normalized by post-multiplication by a diagonal matrix D, giving for the structure on the primary factors

$$F(\Lambda')^{-1}D \quad . \quad . \quad . \quad (65)$$

D is also the matrix of correlations between the reference vectors and the primary factors, for

$$\Lambda'(\Lambda')^{-1}D = D \quad . \quad . \quad . \quad (66)$$

Each primary factor is therefore correlated with its own reference vector but orthogonal to all the others, as can also be easily seen geometrically.

The matrix of intercorrelations of the primary factors is $D\Lambda^{-1}(\Lambda')^{-1}D$ from equation (65).

If W is the pattern on the primary factors p, so that

$$\text{test scores } s = Wp$$

then the structure on the primary factors is also

$$sp' = Wpp'$$

where pp' is the matrix of correlations between the primary factors, and therefore

$$\text{primary factor structure} = WD\Lambda^{-1}(\Lambda')^{-1}D \quad . \quad . \quad (67)$$

$$\text{Also, this structure} \quad = F(\Lambda')^{-1}D \text{ from (65).}$$

Equating these we have:

$$WD\Lambda^{-1} = F$$

$$\text{whence} \quad W = F\Lambda D^{-1} \quad . \quad . \quad . \quad (68)$$

$$= VD^{-1} \quad . \quad . \quad . \quad (69)$$

We have, therefore,

	Structure	Pattern
Reference vectors . .	$F\Lambda$	$F(\Lambda')^{-1}$
Primary factors . .	$F(\Lambda')^{-1}D$	$F\Lambda D^{-1}$

(70)

where the reference-vector pattern has been entered by analogy but could easily be independently found. It will be seen that the structure and pattern of the primary factors are identical with the pattern and structure of the reference vectors except for the diagonal matrix D. The structure of the one is the pattern of the other multiplied by D.

This theorem is not confined to the case of simple structure, but is more general, and applies to any two sets of oblique axes with the same origin O, of which the axes of the one set are intersections of " primes " taken $r - 1$ at a time in the space of r dimensions, and the axes of the other set are lines perpendicular to those primes. By prime is meant a space of one dimension less than the whole, i.e. Thurstone's hyperplane. The projections of any point P on to the one set of axes are identical with the projections thereon of its oblique co-ordinates on the other set, which sentence is equivalent to the matrix identities (see 70)—

$$F\Lambda = F\Lambda D^{-1} \times D$$

and $$F(\Lambda')^{-1}D = F(\Lambda')^{-1} \times D$$

or $$\left.\begin{matrix}\text{Structure}\\ \text{on one set}\end{matrix}\right\} = \left.\begin{matrix}\text{Pattern on}\\ \text{other set}\end{matrix}\right\} \times \left\{\begin{matrix}\text{Cosines to project it}\\ \text{on to the first set.}\end{matrix}\right.$$

A diagram makes this obvious in the two-dimensional case and gives the key to the situation. A perspective diagram of the three-dimensional case is not very difficult to make and is still more illuminating. The vector (or test) OP is the " resultant " of its oblique co-ordinates (the pattern), but not of its projections (the structure). It is of interest to notice that, either on the reference vectors or on the primary factors—

Pattern × Transpose of Structure = Test-correlations.

This serves as a useful check on calculations. It is geometrically immediately obvious. For consider a space defined by n oblique axes, with origin O, and any two points P and Q each at unit distance from O. The directions OP and OQ mav be taken as vectors corresponding to two tests, and cos POQ to the test correlation.

Consider the pattern, on these axes, of OP, and the structure, on the same axes, of OQ. The former is composed of the oblique co-ordinates of the point P, the latter of the projections on the axes of the point Q, which projections (OQ being unity) are cosines. Then the inner product of those oblique co-ordinates of P with these cosines obviously adds up to the projection of OP on OQ, that is to cos POQ, or the correlation coefficient.

In estimating oblique factors by regression, since the *correlations* between factors and tests must be used, the relevant equation is

$$\hat{f}_0 = \{F_0(\Lambda')^{-1}D\}'R^{-1}z \quad . \quad . \quad (70.1)$$

Ledermann's short cut (section 10a above) requires considerable modification for oblique factors. We no longer have

$$R = M_0M_0' + M_1^2 \quad . \quad . \quad . \quad (10)$$

but

$$\text{Pattern} \times \text{transpose of structure} + M_1^2 = R$$

i.e. in Thurstone's notation

$$(F_0\Lambda D^{-1})\{F_0(\Lambda')^{-1}D\}' + F_1^2 = R \quad . \quad (70.2)$$

and using this (Thomson, 1949), we reach the equation

$$\hat{f}_0 = (I + J)^{-1}\{F_0(\Lambda')^{-1}D\}'F_1^{-2}z \quad . \quad . \quad (70.3)$$

where now

$$J = \{F_0(\Lambda')^{-1}D\}'F_1^{-2}(F_0\Lambda D^{-1}) \quad . \quad . \quad (70.4)$$

in place of Ledermann's $J = M_0'M_1^{-2}M_0$.

Only reciprocals of matrices of order equal to the number of common factors are now required, but the calculation, like all concerning oblique factors, is still one of considerable labour.

19a. *Second-order factors.*—The above primary factors can themselves in their turn be factorized into one, two, or more second-order factors, and a factor-specific for each primary. If the rank of the matrix of intercorrelations of the primaries can be reduced by diagonal entries to say two, then the r primaries will be replaced by $r + 2$ second-order factors which will no longer be in the original common-factor space. The correlations of the primaries with these second-order factors will form an oblong matrix with its first two columns filled, but each succeeding column will have only one entry corresponding to a factor-specific, thus :

$$\begin{bmatrix} r & r & r & . & . & . & . \\ r & r & . & r & . & . & . \\ r & r & . & . & r & . & . \\ r & r & . & . & . & r & . \\ r & r & . & . & . & . & r \end{bmatrix} = E \text{ (say)},$$

where subscripts must be supplied to indicate the primary (the row) and the second-order factor (the column).

The primary factors can be thought of as added to the actual tests, their direction cosines being added as rows below F, which thus becomes :

$$\begin{bmatrix} F \\ \hdashline D\Lambda^{-1} \end{bmatrix}$$

Imagine this matrix post-multiplied by a rotating matrix Ψ, with r rows and $r+2$ columns, which will give the correlations with the $r+2$ second-order factors. The lower part of the resulting matrix will be E, which we already know. That is—

$$D\Lambda^{-1}\Psi = E \quad . \quad . \quad . \quad (71)$$

$$\Psi = \Lambda D^{-1}E \quad . \quad . \quad . \quad (72)$$

and the correlations of the original tests with the second-order factors are then :

$$G = F\Psi = F\Lambda D^{-1}E = VD^{-1}E \quad . \quad (73)$$

G is both a structure and a pattern, with continuous columns equal in number to the general second-order factors, followed by a number of columns equal to the number of primaries, this second part forming an orthogonal simple structure.

20. *Boundary conditions.*—These refer to the conditions under which a matrix of correlation coefficients can be explained by orthogonal factors which run each through only a given number of tests. The problem was first raised by Thomson (1919*b*) and a beginning made with its solution (J. R. Thompson, Appendix to Thomson's paper). Various papers by J. R. Thompson culminated in that of 1929, and see also Black (1929). Thomson returned to the problem in connexion with rotations in the

common-factor space (Thomson, 1936*b*), and Ledermann gave rigorous proofs of the theorems enunciated by Thomson and Thompson and extended them (Ledermann, 1936). A *necessary* condition is that if the largest latent root of the matrix of correlations exceeds the integer s, then factors which run through s tests only and have zero loadings in the other tests are certainly inadequate. This rule has not been proved to be *sufficient*, and when applied to the common-factor space only it is certainly not sufficient, though it seems to be a good guide. Ledermann (1936, 170–4) has given a stringent condition as follows. If we define the nullity of a square matrix as order minus rank, then if it is to be possible to factorize orthogonally a matrix of R rank r in such a way that the matrix of loadings contains at least r zeros in each of its columns, the sum of the nullities of all the r-rowed principal minors of R must at least be equal to r.

21. *The sampling of bonds.*—The root idea is that of the *complete* family of variates that can be made by all possible additive combinations of bonds from a given pool, and the *complete* family of correlation coefficients between pairs of these. Thomson (1927*b*) mooted the idea and worked out the example quoted in Chapter XX. He had earlier (1927*a*) showed that with all-or-none bonds the *most probable* value of a correlation coefficient is $\sqrt{(p_1p_2)}$, where the p's are fractions of the whole pool forming the variates, and the *most probable* value of a tetrad-difference F, zero. Mackie (1928*a*) showed that the *mean* tetrad-difference is zero, and its variance, for F_1—

$$\sigma_F^2 = \frac{1}{N-1}\left\{p_1p_3 + p_2p_4 + p_1p_4 + p_2p_3 - 2(p_1p_2p_3 + p_1p_2p_4 + p_1p_3p_4 + p_2p_3p_4) + 4p_1p_2p_3p_4 + \frac{2(N-2)}{(N-1)^2}(1-p_1)(1-p_2)(1-p_3)(1-p_4)\right\}$$

where N is the number of bonds in the whole pool. He found for the *mean* value of r_{12} the value $\sqrt{(p_1p_2)}$, and for its variance—

$$\sigma^2_{r_{12}} = \frac{(1-p_1)(1-p_2)}{N-1}$$

This is not the variance of all possible correlation coefficients, but of those formed by taking fractions p_1 and p_2 of the pool. The whole family of correlation coefficients will be widely scattered by reason of the different values of p, " rich " tests having high correlations, and those with low p, low correlations. Mackie (1929) next extended these formulæ to variable coefficients (i.e. bonds which no longer were all-or-none). He again found the mean value of F to be zero, and for its variance—

$$\sigma_F^2 = \frac{4(N-1)(N-2)}{N^3}\left\{\frac{2}{\pi}\left(1-\frac{2}{\pi}\right)\right\}^2 + \frac{2(N-1)}{N^3}\left\{1-\left(\frac{2}{\pi}\right)^2\right\}^2$$

The presence of $\frac{2}{\pi}$ in this is due to Mackie's limitation to positive loadings of the bonds. Thomson (1935*b*, 72) removed this limitation and found—

$$\sigma_F^2 = \frac{2(N-1)}{N^3}$$

Similarly, Mackie found for variable positive loadings (1929)—

$$\sigma_r^2 = \frac{1}{N}\left\{1-\left(\frac{2}{\pi}\right)^2\right\}$$

and for all loadings Thomson found (1935*b*)—

$$\sigma_r^2 = \frac{1}{N}$$

Thomson suggested without proof that in general, when limits are set to the variability of the loadings of the bonds, resulting in a family of correlation coefficients averaging $\bar{r}$, these correlations will form a distribution with variance—

$$\sigma_r^2 = \frac{1}{N}(1-\bar{r}^2)$$

and will give tetrad-differences averaging zero with a variance—

$$\sigma_t^2 = \frac{4(N-1)(N-2)}{N^3}\left\{\bar{r}(1-\bar{r})\right\}^2 + \frac{2(N-1)}{N^3}(1-\bar{r}^2)^2$$

Summing up, Thomson says (1935*b*, 77–8): "The sampling principle taken alone gives correlations of all values . . . and zero tetrad-differences if N be large. Fitting the sampled elements with weights . . . if the weights may be *any* weights . . . destroys correlation when N is infinite. This means that on the Sampling Theory a certain approximation to 'all-or-none-ness' is a necessary assumption —not to explain zero tetrad-differences, but to explain the existence of correlations of . . . large size. . . . The most important point in all this appears to me to be the fact that *on all these hypotheses the tetrad-differences tend to vanish.* This tendency appears to be a natural one among correlation coefficients."

A tendency for tetrad-differences to vanish means, of course, a still stronger tendency for large minors of the correlational matrix to vanish. In more general terms, therefore, Thomson's theorem is that in a *complete* family of correlation coefficients the rank of the correlation matrix tends towards unity, and that a *random* sample of variates from this family will (in less strong measure) show the same tendency.

REFERENCES

THIS list is not a bibliography, and makes no pretensions to completeness. It has, on the contrary, been kept as short as possible, and in any case contains hardly any mention of experimental articles. Other references will be found in the works here listed.

References to this list in the text are given thus (Mackie, 1929, 17), or, where more than one article by the same author comes in the same year (Burt, 1937*b*, 84). Throughout the text, however, the two important books by Spearman and by Thurstone are referred to by the short titles *Abilities* and *Vectors* respectively, and Thurstone's later book, *Multiple Factor Analysis*, by the abbreviation *M. F. Anal.* Other abbreviations are :

A.J.P. = American Journal of Psychology.
B.J.P. = British Journal of Psychology, General Section.
B.J.P.Statist. = British Journal of Psychology, Statistical Section.
B.J.E.P. = British Journal of Educational Psychology.
J.E.P. = Journal of Educational Psychology.
Pmka. = Psychometrika.

AITKEN, A. C., 1934, "Note on Selection from a Multi-variate Normal Population," *Proc. Edinburgh Math. Soc.*, **4**, 106–10.
1937*a*, "The Evaluation of a Certain Triple-product Matrix," *Proc. Roy. Soc. Edinburgh*, **57**, 172–81.
1937*b*, "The Evaluation of the Latent Roots and Vectors of a Matrix," *ibid.*, **57**, 269–304.

ALEXANDER, W. P., 1935, "Intelligence, Concrete and Abstract," *B.J.P. Monograph Supplement* **19**.

ANASTASI, Anne, 1936, "The Influence of Specific Experience upon Mental Organization," *Genetic Psychol. Monographs* **8**, 245–355.
and Garrett (see under Garrett).

BARTLETT, M. S., 1935, "The Statistical Estimation of *G*," *B.J.P.*, **26**, 199–206.
1937*a*, "The Statistical Conception of Mental Factors," *ibid.*, **28**, 97–104.
1937*b*, "The Development of Correlations among Genetic Components of Ability," *Annals of Eugenics*, **7**, 299–302.
1938, "Methods of estimating Mental Factors," *Nature*, **141**, 609–10.
1948, "Internal and External Factor Analysis," *B.J.P.Statist.*, **1**, 73–81.
1950, "Tests of Significance in Factor Analysis," *B.J.P.Statist.*, **3**, 77–85.

Black, T. P., 1929, "The Probable Error of Some Boundary Conditions, etc." *Proc. Roy. Soc. Edinburgh*, **49**, 72–7.

Blakey, R., 1940, "Re-analysis of a Test of the Theory of Two Factors," *Pmka.*, **5**, 121–36.

Brown, W., 1910, "The Correlation of Mental Abilities," *B.J.P.*, **3**, 296–322.

and Stephenson, W., 1933, "A Test of the Theory of Two Factors," *ibid.*, **23**, 352–70.

1911, and with Thomson, G. H., 1921, 1925, and 1940, *The Essentials of Mental Measurement* (Cambridge).

Burt, C., 1917, *The Distribution and Relations of Educational Abilities* (London).

1936, "The Analysis of Examination Marks," a memorandum in *The Marks of Examiners* (London), by Hartog and Rhodes.

1937*a*, "Methods of Factor Analysis with and without Successive Approximation," *B.J.E.P.*, **7**, 172–95.

1937*b*, "Correlations between Persons," *B.J.P.*, **28**, 59–96.

1938*a*, "The Analysis of Temperament," *B. J. Medical P.*, **17**, 158–88.

1938*b*, "The Unit Hierarchy," *Pmka.*, **3**, 151–68.

1939, "Factorial Analysis. Lines of Possible Reconcilement," *B.J.P.*, **30**, 84–93.

1940, *The Factors of the Mind* (London).

and Stephenson, W., "Alternative Views on Correlations between Persons," *Pmka.*, **4**, 269–82.

Cattell, R. B., 1944*a*, "Cluster Search Methods," *Pmka.*, **9**, 169–84.

1944*b*, "Parallel Proportional Profiles," *Pmka.*, **9**, 267–83.

1945, "The Principal Trait Clusters for Describing Personality," *Psychol. Bull.*, **42**, 129–61.

1946, *Description and Measurement of Personality* (New York).

1948, "Personality Factors in Women," *B.J.P.Statist.*, **1**, 114–130.

Coombs, C. H., 1941, "Criterion for Significant Common Factor Variance," *Pmka.*, **6**, 267–72.

Davey, Constance M., 1926, "A Comparison of Group, Verbal, and Pictorial Intelligence Tests," *B.J.P.*, **17**, 27–48.

Davis, F. B., 1945, "Reliability of Component Scores," *Pmka.*, **10**, 57–60.

Dodd, S. C., 1929, "The Sampling Theory of Intelligence," *B.J.P.*, **19**, 306–27.

Emmett, W. G., 1936, "Sampling Error and the Two-factor Theory," *B.J.P.*, **26**, 362–87.

1949, "Factor Analysis by Lawley's Method of Maximum Likelihood," *B.J.P.Statist.*, II, (2), 90–7. And see page 282.

Ferguson, G. A., 1941, "The Factorial Interpretation of Test Difficulty," *Pmka.*, **6**, 323–29.

Fisher, R. A., 1921, "Mathematical Foundations of Statistics," *Phil. Trans. Roy. Soc.* A222.

FISHER, R. A., 1922, "Goodness of Fit of Regression Formulæ," *Journ. Roy. Stat. Soc.*, **85,** 597–612.
1925, "'Student's' Distribution," *Metron*, **5,** (3), 3–17.
1925, *Statistical Methods for Research Workers* (Edinburgh).
1935, *The Design of Experiments* (Edinburgh).
and Yates, F.,
1938, *Statistical Tables* (Edinburgh).

GARNETT, J. C. M., 1919*a*, "On Certain Independent Factors in Mental Measurement," *Proc. Roy. Soc.*, A, **96,** 91–111.
1919*b*, "General Ability, Cleverness, and Purpose," *B.J.P.*, **9,** 345–66.

GARRETT, H. E., and Anastasi, Anne, 1932, "The Tetrad-difference Criterion and the Measurement of Mental Traits," *Annals New York Acad. Sciences*, **33,** 233–82.

GREEN, B. F., 1950, "Calculation of Weights for Maximum Battery Reliability," *Pmka.*, **15,** 57–61.

HEYWOOD, H. B., 1931, "On Finite Sequences of Real Numbers," *Proc. Roy. Soc.*, A, **134,** 486–501.

HOLZINGER, K. J., 1935, *Preliminary Reports on Spearman-Holzinger Unitary Trait Study*, No. 5; *Introduction to Bifactor Theory* (Chicago).
1937*a*, *Student Manual of Factor Analysis* (Chicago) (assisted by Frances Swineford and Harry Harman).
1940, "A Synthetic Approach to Factor Analysis," *Pmka.*, **5,** 235–50.
1944, "A Simple Method of Factor Analysis," *Pmka.*, **9,** 257–62.
and Harman, H. H., 1937*b*, "Relationships between Factors obtained from Certain Analyses," *J.E.P.*, **28,** 321–45.
and Harman, H. H., 1938, "Comparison of Two Factorial Analyses," *Pmka.*, **3,** 45–60.
and Harman, H. H., 1941, *Factor Analysis* (Chicago).
and Spearman (see under Spearman).

HORST, P., 1941, "A Non-graphical Method for Transforming into Simple Structure," *Pmka.*, **6,** 79–100.
and Smith, S., 1950, "Discrimination of Two Racial Samples," *Pmka.*, **15,** 271–89.

HOTELLING, H., 1933, "Analysis of a Complex of Statistical Variables into Principal Components," *J.E.P.*, **24,** 417–41 and 498–520.
1935*a*, "The Most Predictable Criterion," *J.E.P.*, **26,** 139–42.
1935*b*, "Simplified Calculation of Principal Components," *Pmka.*, **1,** 27–35.
1936, "Relations between Two Sets of Variates," *Biometrika*, **28,** 321–77.

IRWIN, J. O., 1932, "On the Uniqueness of the Factor *g* for General Intelligence," *B.J.P.*, **22,** 359–63.
1933, "A Critical Discussion of the Single-factor Theory," *ibid.*, **23,** 371–81.

KELLEY, T. L., 1923, *Statistical Method* (New York).
1928, *Crossroads in the Mind of Man* (Stanford and Oxford).
1935, *Essential Traits of Mental Life* (Harvard).

LANDAHL, H. D., 1938, "Centroid Orthogonal Transformations," *Pmka.*, **3**, 219–23.

LAWLEY, D. N., 1940, "The Estimation of Factor Loadings by the Method of Maximum Likelihood," *Proc. Roy. Soc. Edinburgh*, **60**, 64–82.
1941, "Further Investigation in Factor Estimation," *ibid.*, **16**, 176–85.
1943*a*, "On Problems Connected with Item Selection and Test Construction," *ibid.*, **61**, 273–87.
1943*b*, "The Application of the Maximum Likelihood Method to Factor Analysis," *B.J.P.*, **33**, 172–175.
1943*c*, "A Note on Karl Pearson's Selection Formulæ," *Proc. Roy. Soc. Edinburgh*, **62**, 28–30.
1944, "The Factorial Analysis of Multiple Item Tests," *ibid.*, **62**, 74–82.
1949, "Problems in Factor Analysis," *ibid.*, **62**, Part IV, (No. 41).

LEDERMANN, W., 1936, "Mathematical Remarks concerning Boundary Conditions in Factorial Analysis," *Pmka.*, **1**, 165–74.
1937*a*, "On the Rank of the Reduced Correlational Matrix in Multiple-factor Analysis," *ibid.*, **2**, 85–93.
1937*b*, "On an Upper Limit for the Latent Roots of a Certain Class of Matrices," *J. Lond. Math. Soc.*, **12**, 14–18.
1938*a*, "A Shortened Method of Estimation of Mental Factors by Regression," *Nature*, **141**, 650.
1938*b*, "Note on Professor Godfrey Thomson's Article on the Influence of Univariate Selection on Factorial Analysis," *B.J.P.*, **29**, 69–73.
1938*c*, "The Orthogonal Transformations of a Factorial Matrix into Itself," *Pmka.*, **3**, 181–87.
1939*a*, "Sampling Distribution and Selection in a Normal Population," *Biometrika*, **30**, 295–304.
1939*b*, "A Shortened Method of Estimation of Mental Factors by Regression," *Pmka.*, **4**, 109–16.
1940, "A Problem Concerning Matrices with Variable Diagonal Elements," *Proc. Roy. Soc. Edinburgh*, **60**, 1–17.
and Thomson (see under Thomson).

LINDQUIST, E. F., 1940, *Statistical Analysis in Education* (Boston).

MACKIE, J., 1928*a*, "The Probable Value of the Tetrad-difference on the Sampling Theory," *B.J.P.*, **19**, 65–76.
1928*b*, "The Sampling Theory as a Variant of the Two-factor Theory," *J.E.P.*, **19**, 614–21.
1929, "Mathematical Consequences of Certain Theories of Mental Ability," *Proc. Roy. Soc. Edinburgh*, **49**, 16–37.

McNemar, Q., 1941, "On the Sampling Errors of Factor Loadings," *Pmka.*, **6**, 141–52.

1942, "On the Number of Factors," *ibid.*, **7**, 9–18.

Medland, F. F., 1947, "An empirical comparison of Methods of Communality Estimation," *Pmka.*, **12**, 101–10.

Mosier, C. I., 1939, "Influence of Chance Error on Simple Structure," *Pmka.*, **4**, 33–44.

Pearson, K., 1902, "On the Influence of Natural Selection on the Variability and Correlation of Organs," *Phil. Trans. Roy. Soc. London*, **200**, A, 1–66.

1912, "On the General Theory of the Influence of Selection on Correlation and Variation," *Biometrika*, **8**, 437–43.

and Filon, L. N. G., 1898, "On the Probable Errors of Frequency Constants and on the Influence of Random Selection on Variation and Correlation," *Phil. Trans. Roy. Soc. London*, **191**, A, 229–311.

and Moul, M., 1927, "The Sampling Errors in the Theory of a Generalized Factor," *Biometrika*, **19**, 246–91.

Jeffery, G. B., and Elderton, E. M., 1929, "On the Distribution of the First Product-moment Coefficient, etc.," *ibid.*, **21**, 191–2.

Peel, E. A., 1947, "A short method for calculating Maximum Battery Reliability," *Nature*, **159**, 816.

1948, "Prediction of a Complex Criterion and Battery Reliability," *B.J.P.Statist.*, **1**, 84–94.

Price, B., 1936, "Homogamy and the Intercorrelation of Capacity Traits," *Annals of Eugenics*, **7**, 22–7.

Reyburn, H. A., and Taylor, J. G., 1939, "Some Factors of Personality," *B.J.P.*, **30**, 151–65.

1941*a*, "Some Factors of Intelligence," *ibid.*, **31**, 249–61.

1941*b*, "Factors in Introversion and Extraversion," *ibid.*, **31**, 335–40.

1943*a*, "On the Interpretation of Common Factors: a Criticism and a Statement," *Pmka.*, **8**, 53–64.

1943*b*, "Some Factors of Temperament: a Re-examination," *ibid.*, **8**, 91–104.

and Raath, M. J., 1949, "Simple Structure: A Critical Examination," *B.J.P.Statist.*, **2**, 125–33.

Spearman, C., 1904, "General Intelligence objectively Determined and Measured," *A.J.P.*, **15**, 201–93.

1913, "Correlation of Sums or Differences," *B.J.P.*, **5**, 417–26.

1914, "The Theory of Two Factors," *Psychol. Rev.*, **21**, 101–15.

1926 and 1932, *The Abilities of Man* (London).

1928, "The Substructure of the Mind," *B.J.P.*, **18**, 249–61.

1931, "Sampling Error of Tetrad-differences," *J.E.P.*, **22**, 388.

1939*a*, "Thurstone's Work Reworked," *J.E.P.*, **30**, 1–16.

1939*b*, "Determination of Factors," *B.J.P.*, **30**, 78–83.

Spearman, C., and Hart, B., 1912, "General Ability, its Existence and Nature," *B.J.P.*, **5**, 51–84.

and Holzinger, K. J., 1924, "The Sampling Error in the Theory of Two Factors," *B.J.P.*, **15**, 17–19.

and Holzinger, K. J., 1925, "Note on the Sampling Error of Tetrad-differences," *ibid.*, **16**, 86–8.

and Holzinger, K. J., 1929, "Average Value for the Probable Error of Tetrad-differences," *ibid.*, **20**, 368–70.

Stephenson, W., 1931, "Tetrad-differences for Verbal Sub-tests relative to Non-verbal Sub-tests," *J.E.P.*, **22**, 334–50.

1935, "Correlating Persons instead of Tests," *Character and Personality*, **4**, 17–24.

1936*a*, "A new Application of Correlation to Averages," *B.J.E.P.*, **6**, 43–57.

1936*b*, "The Inverted-factor Technique," *B.J.P.*, **26**, 344–61.

1936*c*, "Introduction to Inverted-factor Analysis, with some Applications to Studies in Orexis," *J.E.P.*, **27**, 353–67.

1936*d*, "The Foundations of Psychometry: Four Factor Systems," *Pmka.*, **1**, 195–209.

1939, "Abilities Defined as Non-fractional Factors," *B.J.P.*, **30**, 94–104.

and Brown (see under Brown).

and Burt, C. (see under Burt).

Swineford, Frances, 1941, "Comparisons of the Multiple-factor and Bifactor Methods of Analysis," *Pmka.*, **6**, 375–82 (see also Holzinger, 1937*a*).

Thompson, J. R. (see under Thomson, 1919*b*).

1929, "The General Expression for Boundary Conditions and the Limits of Correlation," *Proc. Roy. Soc. Edinburgh*, **49**, 65–71.

Thomson, G. H., 1916, "A Hierarchy without a General Factor," *B.J.P.*, **8**, 271–81.

1919*a*, "On the Cause of Hierarchical Order among Correlation Coefficients," *Proc. Roy. Soc.*, A, **95**, 400–8.

1919*b*, "The Proof or Disproof of the Existence of General Ability" (with Appendix by J. R. Thompson), *B.J.P.*, **9**, 321–36.

1927*a*, "The Tetrad-difference Criterion," *B.J.P.*, **17**, 235–55.

1927*b*, "A Worked-out Example of the Possible Linkages of Four Correlated Variables on the Sampling Theory," *ibid.*, **18**, 68–76.

1934*a*, "Hotelling's Method modified to give Spearman's *g*," *J.E.P.*, **25**, 366–74.

1934*b*, "On measuring *g* and *s* by Tests which break the *g*-hierarchy," *ibid.*, **25**, 204–10.

1935*a*, "The Definition and Measurement of *g* (General Intelligence)," *J.E.P.*, **26**, 241–62.

1935*b*, "On Complete Families of Correlation Coefficients and their Tendency to Zero Tetrad-differences: including a Statement of the Sampling Theory of Abilities," *B.J.P.*, **26**, 63–92.

THOMSON, G. H., 1936*a*, " Some Points of Mathematical Technique in the Factorial Analysis of Ability," *J.E.P.*, **27**, 37–54.
1936*b*, " Boundary Conditions in the Common-factor Space, in the Factorial Analysis of Ability," *Pmka.*, **1**, 155–63.
1937, " Selection and Mental Factors," *Nature*, **140**, 934.
1938*a*, " Methods of estimating Mental Factors," *ibid.*, **141**, 246.
1938*b*, " The Influence of Univariate Selection on the Factorial Analysis of Ability," *B.J.P.*, **28**, 451–9.
1938*c*, " On Maximizing the Specific Factors in the Analysis of Ability," *B.J.E.P.*, **8**, 255–63.
1938*d*, " The Application of Quantitative Methods in Psychology," see *Proc. Roy. Soc. B.*, **125**, 415–34.
1938*e*, " Recent Developments of Statistical Methods in Psychology," *Occupational Psychology*, **12**, 319–25.
1939*a*, " Factorial Analysis. The Present Position and the Problems Confronting Us," *B.J.P.*, **30**, 71–77. " Agreement and Disagreement in Factor Analysis. A Summing Up," *ibid.*, 105–8.
1939*b*, " Natural Variances of Mental Tests, and the Symmetry Criterion," *Nature*, **144**, 516.
1940, " Weighting for Battery Reliability and Prediction," *B.J.P.*, **30**, 357–66.
1943, " A Note on Karl Pearson's Selection Formulæ," *Math. Gazette*, December, 197–8.
1944, " The Applicability of Karl Pearson's Selection Formulæ in Follow-up Experiments," *B.J.P.*, **34**, 105.
1947, " Maximum Correlation of Two Weighted Batteries," *B.J.P. Statist.*, **1**, (1), 27–34.
1948, " Relations of Two Weighted Batteries," *ibid.*, **1**, (2), 82–3.
1949, " On Estimating Oblique Factors," *ibid.*, **2**, (1), 1–2.
and Bailes, S., 1926, " The Reliability of Essay Marks," *The Forum of Education*, **4**, 85–91.
and Brown, W. (see under Brown).
and Ledermann, W., 1938, " The Influence of Multi-variate Selection on the Factorial Analysis of Ability," *B.J.P.*, **29**, 288–306.
THURSTONE, L. L., 1932, *The Theory of Multiple Factors* (Chicago).
1933, *A Simplified Multiple-factor Method* (Chicago).
1935, *The Vectors of Mind* (Chicago).
1938*a*, " Primary Mental Abilities," *Psychometric Monograph No.* 1 (Chicago).
1938*b*, " The Perceptual Factor," *Pmka.*, **3**, 1–18.
1938*c*, " A New Rotational Method," *ibid.*, **3**, 199–218.
1940*a*, " An Experimental Study of Simple Structure," *ibid.*, **5**, 153–68.
1940*b*, " Current Issues in Factor Analysis," *Psychol. Bull.*, **37**, 189–236.
1944*a*, " Second-order Factors," *Pmka.*, **9**, 71–100.

THURSTONE, L. L., 1944*b*, *A Factorial Study of Perception* (Chicago).
1947, *Multiple Factor Analysis* (Chicago).
and Thurstone, T. G., 1941, "Factorial Studies of Intelligence," *Psychometric Monograph*, **2** (Chicago).
THURSTONE, T. G., 1941, "Primary Mental Abilities of Children," *Educ. and Psychol. Meas.*, **1**, 105–16.
TRYON, R. C., 1932*a*, "Multiple Factors *vs* Two Factors as Determiners of Abilities," *Psychol. Rev.*, **39**, 324–51.
1932*b*, "So-called Group Factors as Determiners of Ability," *ibid.*, **39**, 403–39.
1935, "A Theory of Psychological Components—an Alternative to Mathematical Factors," *ibid.*, **42**, 425–54.
1939, *Cluster Analysis* (Berkeley, Cal.).
TUCKER, L. R., 1940, "The Rôle of Correlated Factors in Factor Analysis," *Pmka.*, **5**, 141–52.
1944, "A Semianalytical Method of Rotation to Simple Structure," *Pmka.*, **9**, 43–68.
VERNON, P. E., 1947, "Psychological Tests in the Royal Navy, Army, and A.T.S.," *Occup. Psychol.*, **21**, 53–74.
WHERRY, R. J., 1931, "New Formula for Predicting Shrinkage of Multiple Correlation," *Annals Math. Stat.*, **2**, 440–58.
1949, "New Iterative Method for Correcting Communalities," *Pmka.*, **14**, 231–41.
WILSON, E. B., 1928*a*, "On Hierarchical Correlation Systems," *Proc. Nat. Acad. Sc.*, **14**, 283–91.
1928*b*, "Review of *The Abilities of Man*," *Science*, **67**, 244–8.
1933*a*, "On the Invariance of General Intelligence," *Proc. Nat. Acad. Sc.*, **19**, 768–72.
1933*b*, "Transformations preserving the Tetrad Equations," *ibid.*, **19**, 882–4.
1933*c*, "On Overlap," *ibid.*, **19**, 1039–44.
1934, "On Resolution into Generals and Specifics," *ibid.*, **20**, 193–6.
and Worcester, Jane, 1934, "The Resolution of Four Tests," *ibid.*, **20**, 189–92.
and Worcester, Jane, 1939, "Note on Factor Analysis," *Pmka.*, **4**, 133–48.
WISHART, J., 1928, "Sampling Errors in the Theory of Two Factors," *B.J.P.*, **19**, 180–7.
WORCESTER, Jane, and Wilson (see under Wilson).
WRIGHT, S., 1921, "Systems of Mating," *Genetics*, **6**, 111–78.
YATES, F., (see Fisher and Yates).
YOUNG, G., 1939, "Factor Analysis and the Index of Clustering," *Pmka.*, **4**, 201–8.
1941, "Maximum Likelihood Estimation and Factor Analysis," *ibid.*, **6**, 49–53.
and Householder, A. S., 1940, "Factorial Invariance and Significance," *ibid.*, **5**, 47–56.

INDEX

The numbers refer to pages. (My warm thanks are due to Mr. W. G. Emmett and Mr. A. E. G. Pilliner for making this index.)