THE BOOK WAS DRENCHED

UNIVERSAL
LIBRARY
OU_164417
UNIVERSAL
LIBRARY

OSMANIA UNIVERSITY LIBRARY

Call No. 150
R97F

Acc. No. 2174

Osmania University Library

Call No. 150 B97F Accession No. 21741

Author Burt, C

Title Factors of the Mind.

This book should be returned on or before the date last marked below.

THE FACTORS OF THE MIND

The
FACTORIAL ANALYSIS OF HUMAN ABILITY
by
GODFREY H. THOMSON, D.Sc., PH.D.
16s. *net.*

THE SUB-NORMAL SCHOOL CHILD
by
CYRIL BURT, M.A., D.Sc., LL.D.

VOL. I. THE YOUNG DELINQUENT
Third and Revised Edition
20s. *net.*

VOL. II. THE BACKWARD CHILD
22s. 6d. *net.*

UNIVERSITY OF LONDON PRESS LTD.

THE
FACTORS OF THE MIND

An Introduction to
Factor-analysis in Psychology

BY
CYRIL BURT
M.A., D.Sc. (OXON.), Hon. LL.D. (ABERDEEN)
PROFESSOR OF PSYCHOLOGY IN THE
UNIVERSITY OF LONDON



UNIVERSITY OF LONDON PRESS, LTD.
10 & 11 WARWICK LANE, LONDON, E.C.4

FIRST PRINTED 1940

AGENTS OVERSEAS

AUSTRALIA, NEW ZEALAND AND SOUTH SEA ISLANDS
W. S. SMART, P.O. Box 120 C.C., SYDNEY, N.S.W.

CANADA: CLARKE, IRWIN & Co., Ltd.,
480–486 University Avenue, TORONTO.

INDIA: LONGMANS, GREEN & Co., Ltd.,
BOMBAY: Nicol Road.
CALCUTTA: 17 Chittaranjan Avenue.
MADRAS: 36A Mount Road.

SOUTH AFRICA
H. B TIMMINS, P.O. Box 94, CAPE TOWN.

Printed in Great Britain for the UNIVERSITY OF LONDON PRESS, LTD.,
by HAZELL, WATSON AND VINEY, LTD., London and Aylesbury.

PREFACE

DURING the past ten years, psychologists, both in this country and in America, have shown a rapidly increasing interest in factor-analysis as an instrument of research. In Great Britain, following the remarkable lead of Professor Spearman over thirty-five years ago, a vast stream of factorial work has issued from our laboratories and schools; but a large proportion of it—particularly that which deals, not with results, but with methods—has remained buried in postgraduate theses or in special reports, and so has never become generally known. Indeed, until quite recently, factor-analysis has been looked upon as a peculiar and somewhat isolated branch of psychology—at best a field for specialists, at worst the dubious hobby of an esoteric school, but in any case beyond the ken of the ordinary scientific reader.

Nevertheless, though its abstract foundations are still the subject of some controversy, its concrete applications, particularly in the spheres of education and vocational guidance, have proved more stimulating and more fruitful than any other line of approach. "The entire practice of mental testing and the whole body of individual psychology rest," we have been told, "upon a factorial basis." To-day the theory of mental factors is discussed in almost every psychological textbook. All our students are expected to know something of its leading principles and of its main achievements. And yet there is no work of reference on the subject that does not presuppose a mathematical background that very few students possess. The majority see in it their *pons asinorum*: for factor-analysis is still presented to them as an abstruse statistical technique, too specialized for any but the advanced mathematician to follow or employ.

The standpoint advocated in this book is very different,

I hope to show that, like many other scientific methods that wear a mathematical dress, factor-analysis is merely a refinement of a simple and very ordinary logical procedure. At the same time, I am equally convinced that, in its detailed applications, the logic of a complex subject like psychology cannot be wholly identified with the logic of the simpler sciences. Here perhaps I differ from most other factorists, who explicitly model their initial postulates and the modes of proof on those obtaining in the older sciences, such as physics and astronomy. If, however, the logic of psychology is in some measure peculiar to itself, then it needs studying as a special discipline. In dealing with human problems, whether personal, social, or international, the plain man most frequently goes wrong, not so much because his implicit psychological assumptions are at fault, but rather because his mode of reasoning is inappropriate to psychology. And I hold it to be far more important that the student of a particular science should appreciate the logical method of his science than that he should memorize a mass of details about facts or the latest fashionable theories.

This book, therefore, will be concerned primarily with methods rather than with results. Nevertheless, it has always been for the sake of the results that I have sought to use, adapt, and evaluate the various devices available. As a clinical psychologist in the education department of the London County Council, my interest lay primarily in the discovery of practical tests and in individual diagnosis. My collaborators were teachers and research workers who had often enjoyed a mathematical training. And to them this volume owes an exceptionally heavy debt. Like most of my work, it is based largely on trials and inquiries that they carried out, often without thought of eventual publication. As those who have studied the Council's *Reports* will be aware, nearly all the methods described in the following pages have been tried out in this way upon concrete practical problems, though full details of the procedure have, as a rule, not been published before. The various devices we considered, not as forming a new or a systematic technique, but rather as an *ad hoc* means of verifying impressions or

conclusions indirectly gleaned from first-hand surveys or from the clinical study of individual cases.

On receiving a part-time chair, I found myself engaged in instruction as well as in research. It therefore became necessary to systematize the working methods that had seemed most promising, and to set them out in an intelligible form. In the absence of any textbook, we evolved a series of roneo'd notes, containing simplified proofs, computers' instructions, and a set of worked examples. As the demand for these increased, the University of London Press very generously agreed to publish, as a sequel to my earlier volumes, which had already incorporated results from factorial and other statistical inquiries, a further work explaining in greater detail the technical methods we had used. On the outbreak of war it appeared that I should be separated from most of my research students, and engaged on novel problems, which might still nevertheless demand somewhat similar methods of analysis. Accordingly, I resolved to issue without further delay the substance of my notes and lectures in book form. Later events, not foreseen at the time of this decision, must be in part my apology for the imperfections the reader will no doubt discover.

To Mr. W. Stanley Murrell, Manager of the University of London Press, all psychologists owe a high tribute of gratitude for the enterprise and generosity he has always shown in accepting scientific publications on so young a subject as psychology. I personally have to thank him for the care, the promptitude, and the skill with which he has seen this work through the press at a time of unusual difficulty. I have also to acknowledge the kindness of the London County Council in permitting me to incorporate tables, formulæ, and other matter that had originally appeared in published and unpublished reports drawn up while I was in their service.

The book consists of three main Parts. The first explains the general method of factor-analysis as a logical technique; the second describes the relations between the various forms which this general technique has taken, either for the purpose of special problems or at the hands of parti-

cular writers; the last attempts to illustrate and defend the views I have advanced in what would seem to be the most effective way—namely, by an actual application of the alternative methods to a concrete issue of general interest: the problem of temperamental types.

Part I is itself the outcome of an experiment. Two years ago, I found myself compelled at short notice to take over a course of lectures on factor-analysis to a group of students, who, like most psychological students in this country, had little mathematical or statistical knowledge. Technicalities were therefore reduced to a minimum; and the main argument was based on a logical rather than a mathematical approach. Presented in this light, the fundamental ideas of the statistical factorist became at once clearer and more interesting both to the mathematical and the non-mathematical; at the same time, they illuminated questions in psychology and other sciences, not generall regarded as factorial.

In arranging these lectures for the press I have also kept in mind the needs of the research-worker. I have discussed in some detail the fallacies that invalidate so many theses on the subject; and, since research students enter psychology by such different paths, I have inserted a wide variety of proofs and illustrations, in the hope that one at least may be intelligible to every reader. For the same reason, it seemed necessary to expatiate at greater length on what may be called the present growing-points—particular issues on which in the near future research is likely to concentrate, such as the problem of metric units, of mental energy as a factorial concept, of correlating persons instead of tests, and of the various hierarchical criteria. To avoid breaking the general thread, these side-issues, and the supplementary illustrations and interpolated proofs, have been printed in smaller type. They are sections which the general reader, at any rate on a first perusal, will merely skip or skim.

To the advanced mathematical psychologist, this mode of presentation may seem alike prolix and inconclusive. To forestall this criticism, I had originally intended to append a systematic summary, giving rigorous algebraic proofs of the

essential formulæ, on the lines of my published *Memorandum* drawn up at the request of the International Examinations Inquiry Committee. But the recent publications of Thurstone and Thomson—both remarkable for the lucidity and thoroughness with which they have covered these more technical aspects—have rendered my own attempt not only superfluous but largely out-of-date. Accordingly, I have merely retained an Appendix on working methods which I hope may be of service to the practical investigator, not only in psychology, but also in many other sciences—medicine, agriculture, biology, economics, and the various branches of social science—where, as I believe, factor-analysis might often be applied with ease and advantage.

Parts II and III were actually written before Part I. The second was accepted for publication by the editor of the *British Journal of Psychology* a year or more ago; the third was submitted to the editor of the *British Journal of Medical Psychology* as a sequel to an earlier article printed in its pages. To the editors themselves I am much indebted for permission to include them here instead of in those journals. As they stand, they form a natural continuation of the earlier chapters. The importance of the two issues with which they deal became evident at the symposium on 'Factor-analysis' arranged by the British Psychological Society at the beginning of 1939; and the interest aroused by these special aspects showed the need for a fuller discussion in a more widely accessible form.

As regards the first of these issues, I myself have always held that the different methods of factorizing a given set of data were merely variants of the same underlying principle—alternative ways of reaching the same essential results with a greater or less degree of ease and approximation. This view has encountered strong criticism, and therefore requires more adequate support. Proofs and illustrations are supplied in Part II; and, if they win assent, should do much to reconcile conflicting points of view. Let me add that every student should test the issue for himself. Whatever conclusion he reaches, there can be no better way of understanding general principles than by trying all the

different procedures in turn on some small table of measurements.

As regards the second problem, I have always held that the methods of factor-analysis might be applied quite as legitimately to correlations between persons as to correlations between traits, and that the same factors would be reached by either approach. This I have regarded as almost self-evident: yet it has also become the subject of recent attack. Until an agreement on this issue is achieved, the very nature of mental factors must remain in doubt. In defending my own position, it seemed that the most convincing line was, even at the risk of appearing polemical, to examine one by one the various criticisms advanced, and then, adopting the critic's own procedure, to show how the results achieved are virtually the same. This is the aim of the concluding Part. The upshot, I think, is to demonstrate that for practical work a very much simpler arithmetical technique will suffice than is described in the usual books on the subject.

My acknowledgments are too numerous to record in full. I owe most, I fancy, to the writings of those who are not psychologists at all, but have been concerned primarily with the general methodology of the more complex sciences—particularly Keynes, Johnson, and Fisher, and, among mathematicians, Cullis, Sheppard, Russell, and Weyl. To English psychologists, who have engaged in factorial work—Spearman, William Brown, Godfrey Thomson, and my own students—my indebtedness will be manifest at almost every point. To my teacher, William McDougall, I shall never be able to express my thanks: though his own approach to psychological problems was along very different lines, my first factorial research was carried out over thirty years ago under his immediate supervision. Spearman's pre-eminence is acknowledged by every factorist, even by those who at one time differed from him most strongly. My own obligation is a personal one as well: the generosity that he showed in encouraging and criticizing my early work has continued to the present day. The last two Parts of this volume were completed before the appearance of Thomson's admirable work on *The Factorial Analysis of*

Human Ability; but the first Part has gained greatly by his survey of the general field and his sympathetic criticisms of my views. Among the numerous American workers who have entered the field, to our great advantage, the remarkably clear and systematic exposition of Thurstone has proved most suggestive; he will, I hope, forgive me for taking his own tables as an occasional text for my discussions.

To my recent colleagues, Dr. W. Stephenson and Dr. A. J. Marshall, Research Assistants in the Department of Psychology, this book owes an unusually heavy debt. They have always been ready to read my notes, criticize my views, and even check my calculations. Circumstances have lately deprived me of their help: otherwise I should have held my manuscript back, and profited still more fully from their criticisms. To Stephenson, one of the most original and vigorous of the many students who have worked both under Spearman and myself, I am particularly grateful. Nothing is more stimulating than the presence of an enthusiastic collaborator, eager to explore a new field of work, yet attacking it from an opposite angle instead of along identical lines. On the two main issues I have just mentioned, his outspoken criticisms, and above all the opportunities we have had for personal discussion, have been invaluable at every point. I believe that both the problems at stake and the alternative solutions have been made at once clearer and more interesting to the beginner, because I have thus been able to attack them, not by a dogmatic pronouncement from one side, but as part of a friendly and lively debate.

Throughout, however, my ultimate desire has been to emphasize the agreement rather than the disagreement between the various schools. When the differences are fairly faced, the final outcome, I am convinced, is nearly always to discover that the antagonistic doctrines are by no means so incompatible as has been supposed: each has merely been stressing a complementary aspect of the truth which the opposite approach had missed or underrated. If, therefore, this book succeeds in establishing a unity of principle underlying the mutual criticisms and the divergent views of the various contributors, then one important step will

have been taken towards establishing factor-analysis on a settled basis. When that is achieved, I believe we shall see in it, no longer a special branch of psychological research, but a logical technique available for use in every complex science.

C. B.

University College, London.
September, 1940.

CONTENTS

PAGE

PREFACE V

PART I. THE LOGICAL AND METAPHYSICAL STATUS OF FACTORS IN PSYCHOLOGY

CHAPTER

I. WHY DOES THE PSYCHOLOGIST NEED FACTORS? . 3

II. FACTORS ARE VARIOUSLY USED FOR DESCRIPTION, PREDICTION, CAUSAL EXPLANATION . 14

III. THE GENERAL NATURE OF THE FACTORIAL TECHNIQUE 72

IV. THE LOGICAL STATUS OF MENTAL FACTORS. . 95

V. THE DERIVATION OF THE CHIEF FACTOR THEORIES 139

VI. THE DIFFERENCES BETWEEN P-, Q-, AND R-TECHNIQUES 169

VII. THE METAPHYSICAL STATUS OF MENTAL FACTORS 210

VIII. SUMMARY AND CONCLUSIONS . . . 249

PART II. THE RELATIONS BETWEEN DIFFERENT METHODS OF FACTOR-ANALYSIS

IX. CLASSIFICATION OF METHODS . . . 255

X. VARIANCE, COVARIANCE, AND CORRELATION . 271

XI. CORRELATIONS BETWEEN TESTS AND BETWEEN PERSONS 289

CHAPTER PAGE

XII. GENERAL-FACTOR METHODS AND GROUP-FACTOR METHODS 295

XIII. SIMPLE SUMMATION METHODS AND WEIGHTED SUMMATION METHODS 320

XIV. TESTS OF SIGNIFICANCE AND HIERARCHICAL TENDENCY 333

XV. SUMMARY AND CONCLUSIONS 365

PART III. THE DISTRIBUTION OF TEMPERAMENTAL TYPES

XVI. A REPLY TO CRITICISMS OF THE METHOD . . 371

XVII. ANALYSIS OF AN ILLUSTRATIVE GROUP . . 387

XVIII. A REPLY TO CRITICISMS OF THE RESULTS . . 410

XIX. FINAL CONCLUSIONS 424

APPENDIX I. WORKING METHODS FOR COMPUTERS 447

APPENDIX II. ANALYSIS OF A MATRIX INTO ITS LATENT ROOTS AND VECTORS . . . 487

REFERENCES 495

INDEX OF AUTHORS 503

INDEX OF SUBJECTS 505

PART I

THE LOGICAL AND METAPHYSICAL STATUS OF FACTORS IN PSYCHOLOGY

CHAPTER I

WHY DOES THE PSYCHOLOGIST NEED FACTORS?

The Origin of Factor-analysis.—It is impossible to understand the nature of a mental factor unless we first understand the nature of the technique by which it is derived. Historically what we now call factor-analysis is a mathematical procedure developed by psychologists as an extension of the ordinary device of correlation: "the commencement," says Spearman (to whom more than to anyone else the introduction of factorial methods is due), "consisted in noting that, when any pair of abilities are correlated with each other, they can be regarded *as depending on a common factor*" ([56], p. i).[1] Correlation in its turn, or so it has generally been maintained, is merely a statistical application of what Mill called the 'method of concomitant variations'; and Mill's fifth 'canon' might well be taken to express the implicit assumption on which nearly all interpretations of factor-analysis have been tacitly based: "Whatever phenomenon varies in any manner, whenever another phenomenon varies in some particular manner, is either a cause or an effect of that phenomenon, or is else connected with it *through some fact of causation.*"[2]

In psychology, as in the biological sciences generally, the processes are too involved for us to isolate, as a directly observable 'phenomenon,' either a simple 'cause' or a simple 'effect': at most, we can only surmise that some

[1] Figures in brackets refer to the numbers of the publications given in the list of references at the end of the volume. In the quotation I should prefer to substitute the word 'variables' for the word 'abilities,' since the observation that two sets of correlated measurements might be expressed in terms of a common linear component was not really peculiar to the psychologist.

[2] *System of Logic*, Bk. I, p. 441. Mill, it may be remarked, is practically the only logician whom factorists ever quote in support of their formal procedure.

underlying ' fact of causation ' connects the visible changes we can actually observe. Hence statistical analysis has to be used to supplement or take the place of experimental analysis, in order that we may allow for the complex mass of irrelevant influences which, in the simpler sciences like physics or chemistry, we should usually be able to remove or control. And in general the psychologist is able to state the connexions between the facts observed only in terms of *partial* dependence, seldom if ever in terms of *perfect* dependence. The degree of such partial dependence is measured by a coefficient of correlation or its equivalent; and, as Spearman puts it, " the system of correlation proposed by Galton and elaborated by Pearson . . . may be conceived as expressing the *hidden underlying cause* of the variations investigated."[1] To designate the supposed

[1] Spearman, ' The Proof and Measurement of the Association between Two Things,' *Am. J. Psychol.*, XV, 1904, pp. 74–5 (his italics). Spearman himself would prefer to use " not Galton's measure of correlation, but the square thereof " to " indicate the relative influence of the factors "—an early hint of the two-factor formula. The last phrase seems to contain the first use, at any rate in psychological writings, of the word ' factor ' (or rather ' factors ') in the sense of ' component '—a term which is also employed by Spearman in the same passage. In this article, however, the word designates a cause of positive correlation shared by *two* variables only. The extension of the idea (not of the term) to a cause of correlation between a *number* of variables was made in Spearman's second contribution ([12], 1904), where the idea of a ' Hierarchy of the Specific Intelligences ' was first broached. Following a suggestion of Galton, Cattell (1890–1901) had tested long series of students with laboratory tests, but found ' little inter-correlation between the tests ' and therefore no evidence for any common factor ([6], [9]). This seemed confirmed by Thorndike and others. Spearman, however, holding that students form a group too homogeneous to exhibit wide variations in intelligence, repeated the work on a smaller scale with groups of children. Using Yule's formula for partial correlation [8], and a correction formula of his own, he showed that there is a ' common and essential element ' underlying all the ' intelligences,' which he identified with general sensory discrimination. My own early work was chiefly directed towards showing that the " *highest common factor* was exhibited," not only (and not most clearly) in the simpler sensory processes, such as Cattell and Spearman had tested, but also (and still more clearly) " in the more complex mental processes, especially those demanding logical thought " ([16], 1909; cf. [20]): this entailed a wider variety of tests, and (to cope with the larger correlation tables thus obtained) a modified statistical procedure, which was based on a ' summation method ' and the ' product

'underlying cause'—Mill's 'common connecting fact of causation'—the name 'factor' is now regularly used in psychology.[1]

An example from an early research will make the form of reasoning clear. On testing a group of children for the chief school subjects, it was found that, "as a rule, those who are bad at reading are bad at spelling as well; their arithmetic is also below the average for their age, but by no means as bad as their reading or spelling." Now, we cannot suppose that weakness in spelling is (to borrow Mill's language) 'either a cause or an effect' of weakness in arithmetic. Consequently, we infer that both are 'connected' through some more fundamental cause, which we term a 'common factor.' We suppose, for instance, that an underlying ability—'general intelligence'—is mainly responsible both for progress and for weakness in all three subjects. If spelling and reading further vary together in a way that is not wholly accounted for by the factor common to all three subjects, we apply Mill's 'fourth canon'—the 'method of residues'; we eliminate what is due to the first factor, and decide by a fresh application of the method of concomitant variation whether or not there is yet another factor—verbal facility, for example—common to reading and spelling, but not shared by arithmetic.

The Measurement of Factors.—To render the arguments more precise, we endeavour at every stage to *measure* the amount of 'concomitant variation.' This, of course, means that we must begin by measuring the mental abilities themselves. Standardized tests are employed; and, as a rule, implicitly if not explicitly, the examinees' performances in the tests are first translated into terms of the variability of the group that has been tested, i.e. into terms of their own standard deviation, which is treated as a universal unit

theorem' (cf. below, pp. 150–2). The two-factor theory as such was propounded by Spearman and Hart in 1912 ([24]: cf. also [28]). For the further history of the subject, see Brown and Thomson [39] and [132].

[1] At this stage I shall not attempt any more formal definition of the word 'factor.' Such a definition, I hold, must come at the end, not at the beginning, of our inquiry (cf. p. 256). For the time being we may accept the term as meaning what factorists claim to deduce by analysing their correlations according to some accredited technique (see pp. 210 f.).

of measurement: the resemblances between their performances can then be measured by the average product of all the pairs. This is Galton's 'index of co-relation' as calculated by the so-called product-moment method.

It is important, however, to realize that the preliminary standardization of the unit is not indispensable. All that is necessary for the calculation is that the measurements should be expressed as deviations from their own average. Then, even though the measurements have not been reduced to terms of their own standard deviations, we can still use the average product of the pairs to measure the amount of concomitant variation. The product-moment is then termed the covariance; and in theory all forms of factor-analysis can be applied to tables of covariances just as well as to tables of correlations.

When the calculation is completed, the degree to which each test-performance appears to depend on the fundamental ability or 'factor' supposed to influence it is finally stated in terms of a 'factor loading' or 'factor saturation,' as it is variously called, that is, a similar coefficient or 'index' measuring the amount of resemblance (or 'co-relation') between the empirical test on the one hand, and the estimated measurements for the hypothetical factor on the other. Tests of significance are generally applied, not to the factor loadings or factor saturations, but to the coefficients of correlation or their residuals; and, as usual, they indicate, not the probability that the postulated hypothesis is true, but the probability that the figures tested may after all have arisen from the mere effects of random sampling, that is, from what is loosely called 'chance.'

'*The Factors of the Mind.*'—By applying calculations of this kind to the results of mental tests, psychologists have hoped to reach an inventory of what, in Spearman's phrase, have been described as 'the abilities of man.' Verbal ability, arithmetical ability, mechanical ability, retentivity, quickness, perseveration, oscillation of attention, and above all a general factor of intelligence that enters into all we say or do or think—these, or qualities somewhat like them (for, to avoid misconception, the factorists prefer to designate their factors by letters rather than by concrete names), are

supposed to be the ' primary abilities ' that make up the human mind.[1] Other psychologists have gone on to claim that, not only the intellectual or cognitive aspect, but also the emotional or conative aspect of the mind can be described in terms of definable factors.[2] This branch of the work has attracted less attention. Nevertheless, it would clearly be a mistake to begin by identifying the ' factors of the mind,' as Thurstone and Alexander appear to do, exclusively with cognitive ' abilities.'

The catalogue no doubt is still incomplete ; but, we are assured, the number of ' fundamental traits ' that have eluded discovery must now be very small. " It seems to be a fact," says a leading exponent of the Spearman school, " that there is only a limited number of such fundamental tendencies in the human being : Spearman has found five or so ; Thurstone specifies seven ; the Thorndike Unitary Traits Committee expects to find anything between one and twenty." And the writer concludes : " the implication is that these few fundamental factors account for, explain, or are the cause of, all human conduct " ([96], p. 208).

Still more recently, similar factors have been invoked to explain the resemblances, not only between test-performances and temperamental traits, but also between human individuals—resemblances which tempt us to class them together in groups under the heading of ' mental types.' [3] As before the factors are deduced from sets of correlation coefficients or covariances : only now we start by correlating, not the measurements for two tests, but the measurements for two individuals, taking all possible pairs of persons, just as previously we took all possible pairs of tests. Unlike the ' trait-factors,' these ' type-factors,' it is declared, may be exceedingly numerous. " There are," so we are told, " only a few fundamental tendencies in the human being,

[1] Cf. C. Spearman, *loc. cit.*, esp. pp. 411 *et seq.* A clear and convenient summary of methods and results is to be found in Guilford, *Psychometric Methods*, 1936, chap. xiv, esp. pp. 510 *et seq.*

[2] The first to be discovered were again the ' general ' factors—e.g. general emotionality (Burt, *Brit. Ass. Ann. Rep.*, 1915, pp. 694 *et seq.*) and a general moral factor (Webb, *Brit. J. Psych. Mon. Supp.*, 1915), the latter being subsequently accepted by Spearman (*loc. cit.*, p. 359).

[3] Burt, ' Correlations between Persons ' [101] pp. 59–96.

and therefore only a few unitary traits in the mind; but types exist in great numbers." [1]

It is frequently implied that these 'factors of the mind' are innate factors—fundamental elements in the individual's mental endowment handed on to him at birth. Thus, in one of the earliest investigations on intelligence tests, an attempt was made to show that the factor which they tested was not only general but also inborn ([16], pp. 169 f.). And some of the earlier investigations into type-factors, particularly those that appeared to be associated with temperament, race, or sex, suggested the possibility that the most fundamental of all would be those attributable to genetic elements, obeying Mendelian laws and producing traits either linked or segregating freely.[2] It is, however, somewhat unfortunate that the term 'factor' is used for both conceptions—the statistical factors that we are discussing here and the genetic factors responsible for hereditary resemblances: the common name tempts the lay reader and the student to identify the two.

[1] Stephenson [96], p. 209.

[2] Cf. Burt, 'The Inheritance of Mental Characters' ([22], pp. 188 *et seq.*) and 'The Mental Differences between the Sexes' ([23], pp. 380 *et seq.*). The experimental data reported in these early papers gave, so it seemed to me, a strong support to the notion of inheritable group-factors, at a time when the existence of group-factors was generally doubted by psychological factorists. The most obvious but by no means the only instances appeared to be certain well-marked sensory anomalies, such as colour blindness, for which (at any rate in certain forms) the pedigrees show a distribution very similar to that of hæmophilia and other sex-linked recessives. Red-green vision is obviously not a general factor, nor yet a specific factor peculiar to a single test. Nor could it be explained away by arguing that "most so-called group-factors would appear to be the result of education or experience, and so form only apparent exceptions to the two-factor theory." Similar hints of a possible sex-linkage were found in other forms of sensory capacity, in imagery, memory, the verbal factor, the manual factor, the numerical factor, and certain temperamental tendencies traceable to sex-differences in the endocrine glands—characteristics closely related to recognizable group-factors found in other factorial inquiries. The studies of racial types, combining physical with temperamental traits, also suggested the probability of an ultimate discrete basis. But in all such cases the relation between the 'manifest types' and the 'latent genetic types' appeared extremely indirect and complex, so that for all practical purposes it seemed safer to talk "not of mental types, but rather of mental tendencies" (see below, p. 246).

About the genetic factors that influence mental ability and temperament comparatively little is known as yet. Children undoubtedly resemble their parents in regard to general intelligence and many other mental factors, and that in a degree that cannot be wholly explained by post-natal influences. Yet the relation between the observable phenomena is exceedingly indirect, and typical of the remote and complex type of causal determination with which the correlationist has to deal in psychology. The most that we can say about mental inheritance with any assurance is that each individual apparently receives through his two parents a very large sample of a still larger number of unit-determiners; that this sample is mainly but by no means entirely random (certain groups of determiners, for example, being always carried on the same chromosome); that his subsequent development must involve a further sampling of this sample (or rather of its possible effects); and that his mental reaction in any given situation must depend on yet another process of selecting or sampling whatever tendencies have thus developed or survived. It follows that, with few exceptions, the overt mental types, which are all that the psychologist can detect with his tests and rating scales, are related only in a very remote and indirect fashion to inherited types or tendencies: they are, as the biologist would say, phenotypes rather than genotypes. Indeed, if there were any likelihood of establishing mental genotypes, factor-analysis, I imagine, would hardly be the main line of approach which the genetic psychologist would adopt in his endeavours to discover them.

The Criticism of Factors.—In seeking to demonstrate the existence of the mental factors I have described, different psychologists have employed different modes of calculation. As a consequence, they have reached somewhat discrepant conclusions. Each, therefore, has been tempted to criticize any method yielding results a little different from his own. So far, however, the validity of factor-analysis as such has not been seriously questioned. The non-statistical psychologist, it is true, is always a little dubious of statistical demonstrations; but no systematic refutation of the pro-

cedure as a whole has ever been attempted. In a later chapter[1] I shall show that the results of any one method bear direct and simple relations to those of any other; so that all the methods stand or fall together. The detailed differences between the various devices hitherto proposed have been described and discussed with great impartiality and clearness by Prof. Godfrey Thomson in a recent publication.[2] Hence their special features need not detain us here.

Prof. Thomson's book, however, has brought the chief issues to a head. Though he has never condemned the general method in itself, he has always been one of the most vigorous critics of the conclusions popularly drawn. The aims of factor-analysis, as usually stated, he readily accepts: its objects, he says, are twofold—" 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."[3] But whether the mental factors thus arrived at will be so few or so simple as is commonly maintained, he very much doubts. From the first he has opposed the familiar theory that there is a single central factor pervading and dominating all the activities of the mind—' Spearman's *g* ' or ' general intelligence,' as it is variously termed; and in his more recent discussions he goes farther still, and rejects the whole notion that the human mind may be constructed out of a small number of fundamental capacities or traits. " Far from being divided up into unitary factors, the mind is a rich, comparatively undifferentiated complex of innumerable influences—on the physiological side an intricate network of possibilities of intercommunication." The mathematical peculiarities exhibited by our correlation tables are attributable, so he believes, not to psychological laws, but to statistical laws: they are at bottom simply the result of sampling the innumerable factorial elements of which the mind is ulti-

[1] See below, pp. 365 f.

[2] *The Factorial Analysis of Human Ability*, 1939. Cf. also Appendix I for illustrations of the chief methods.

[3] *Loc. cit.*, p. 3.

mately composed, elements which he apparently would identify with the 'neurone arcs' of which the central nervous system is built up.

It would seem, however, that what Thomson is treating as the ultimate factors are something quite different from what Spearman, Thurstone, Kelley, Alexander, and most other psychologists have had in mind, namely, what they would call the primary intellectual abilities—*g*, *v*, *c*, *F*, and the like. Spearman's 'basic components' are rather like the organs of the body; Thomson's are more like its cells; or (to adopt an analogy which both writers use) Spearman's are like the 'parts' of a motor-car—the wheels, the lamps, the horn, the engine, and the tank containing the petrol; Thomson's, more like the ultimate molecules of which all the materials are composed. And there is this further difference between them: Spearman looks upon the mind as a heterogeneous structure built up out of a few essential mechanisms or components; Thomson insists that the mind is almost devoid of structure—a tissue of homogeneous cells rather than an organized whole of specialized parts.[1]

The Reasons for Factor-analysis.—What may be the ultimate structure of the mind, and whether its parts are numerous or few, and its elements similar or differentiated, are questions, so at least it seems to me, which must be eventually decided by other lines of research—physiological, biological, introspective, and experimental.[2] Our present crude distinctions between intellectual abilities may give

[1] The difference is largely one of emphasis. Thus, although in the passage quoted and again on p. 280 he speaks of the mind as '*comparatively* structureless,' elsewhere (pp. 51 and 283) he describes the mind as divided into 'regions' or 'subpools' and even admits that *part* of this 'structure' may be 'innate,' though most of it is due to environment and education. Thus qualified, his two statements are not incompatible, and are far less in conflict with Spearman's than a first reading might suggest.

[2] How the results of these alternative lines of inquiry are related to those of factor-analysis I have tried briefly to indicate elsewhere (e.g. [128], pp. 191–5); it should be added, however, that not every exponent of mental factors would admit the relevance of physiological or other evidence: Stephenson, for example, has recently argued that my 'endeavour to bring physiology into the picture only confuses the issue.'

way to distinctions between the functions of various cortical areas or cell-layers (indeed, the chief group-factors so far discovered rather suggest some such basis); and our distinctions between temperamental types may be resolved into biochemical differences produced by variations in the balance of endocrine secretions: so that in these directions factor-analysis may turn out to be a mere make-shift—a temporary expedient that we may conveniently exploit while awaiting a more refined experimental technique. Since the field is highly complex, a direct advance by non-statistical methods is bound to be slow. Meanwhile, scientific curiosity demands at least a provisional solution; and the immediate needs of applied psychology call for working hypotheses and some practical device for determining the key-characteristics of different individuals. It is these urgent demands that factor-analysis endeavours to meet. How far can we trust it?

Accordingly, in the opening chapters I shall confine myself chiefly to the question of methodology. I propose first to examine what nearly every writer on factor-analysis has hitherto taken for granted, namely, the nature of factor-analysis in itself as a general method of scientific inquiry, irrespective of the particular arithmetical form which the procedure assumes in the hands of this investigator or that, and regardless of the particular results obtained. Only after we have examined the *logical* nature of the arguments by which factors are derived (or so at least I believe) can we go on to decide what may be the *physical* or *meta-physical* nature of the factors hitherto suggested.

If I am correct, the main reason for the protracted controversies which have obscured the whole subject lies in the fact that the opposing parties, though nominally acknowledging the same general purposes, are interested each almost exclusively in one purpose alone—the theoretical analysis of the human mind in the one case, and the practical prediction of individual progress and development in the other. Spearman's original concern, as the title of his great work implies, lay in the abstract nature of intelligence and cognition; his aim was 'to discover the causal mechanisms of the mind and the general laws

which they obey.' Thomson's starting-point, as he himself has related, was an endeavour to improve the methods of selecting pupils for different types of school and career by scholastic examinations or by mental tests. I myself would rather place the initial emphasis on a third and somewhat lowlier purpose. It is one which, I am sure, both parties would accept as equally obvious, yet at the same time one which, just because it is so easily taken for granted and perhaps because it is less ambitious, has been continually passed by. In my view the *primary* object of factorial methods is neither causal interpretation, nor statistical prediction, but exact and systematic description. And I suspect that most of the confusion has arisen because factors, like the correlation coefficients on which they are based, have been invoked to fulfil these three very different purposes, and so have made their appearance at three very different levels of thought—like the famous legal firm of Arkles, Arkles & Arkles, which, " more to its own satisfaction than that of its clients, canvassed three different lines of business in three small offices on three different floors."

CHAPTER II

FACTORS ARE VARIOUSLY USED FOR DESCRIPTION, PREDICTION, OR CAUSAL EXPLANATION

(I) *Description.*—Primarily the figures that specify the psychologist's factors, like those that specify his correlations, are simply numerical constants, descriptive of the sample he is investigating at the moment. But it will at once be asked, if mere description is his immediate aim, why does he go out of his way to express it in terms of abstract hypothetical concepts, like, *g*, *v*, *p*, and the rest, instead of by the concrete and familiar processes he has actually tested, such as reading, spelling, or arithmetic? The motive is usually[1] said to be economy—'a desire for simplification.' "Entities," we are constantly reminded, "should not be multiplied beyond necessity"; and the use of factors will enable us to cut down the number of our working concepts. I would rather describe the underlying purpose as 'orderly simplification': the effort to economize labour and thought is to my mind secondary to the endeavour to make things intelligible, to 'bring order out of chaos.' The task of science is to *organize* ideas rather than to minimize them.

The analysis of correlations fulfils the same end for the multi-variate universe as the analysis of a frequency-

[1] Cf. Guilford, 1936, *Psychometric Methods*, p. 457. "Science . . . wants to know what is the least possible number of concepts with which one can order and describe the multiplicity of phenomena." "Both practical necessity and the desire for theoretical simplification lead one to seek for a few tests which will describe the individual. . . . Such tests might then be said to measure the factors of the mind" (Thomson, *loc. cit.*, p. 4). "The decision may be made on the ground that we should be parsimonious" (*loc. cit.*, p. 15). Similarly Thurstone asks: "What is the minimum number of factors that will account for the observed intercorrelations?" (*loc. cit.*, p. 150); for he too maintains that "the chief object of science is to minimize mental effort" (p. 45).

distribution fulfils for the uni-variate universe. In dealing with a uni-variate universe of measurements, we are not content merely to give a detailed table of frequencies; we seek a few descriptive constants which will specify the leading characteristics of our sample—the mean, the variance (or s.d.), and perhaps the third and higher moments. So too in dealing with a multi-variate universe: instead of merely giving a table of detailed correlations, we seek a few appropriate constants which will characterize the sample, and indirectly the universe from which the sample has presumably been drawn—namely, the variances for the first, second, and possibly other 'factors,' and the corresponding factor-measurements. Our choice of factorial constants, like our preference for the standard deviation rather than the mean variation, will be determined, not entirely by labour-saving considerations, but by a desire for the most pregnant specifications, i.e. for characteristics containing within themselves the largest number of logical implications. In short, our object is a theoretical as well as a practical economy.

(*a*) The '*practical* economy' is perhaps clearest in the field where most of the work has already been done, namely, in the testing of school children. If, for example, we can justifiably group pupils into schools, forms, classes, or standards, according to the general educational ability of each one, there will be a manifest saving of labour: if, however, such attempts at broader classification prove to be unwarranted by the facts, and if it proves wholly fallacious to speak of children as *generally* backward or as *generally* advanced, if, in fact, we can adequately describe them only by their detailed attainments in each separate subject in turn, then the task of educational classification will become exceedingly complex. Factor-analysis begins by checking the validity of such generalized descriptions.

But a description in terms of general ability forms only the first step in dealing with the individual child. It claims to tell us whether he is, on the whole, more likely to succeed or to fail, and by how much; but we also want to know in greater detail where he is likely to succeed or fail, and why. And so we pass from the general factor to the more specific. Yet even here we still seek to be as general as we can: if,

for example, we can say that James is weak in all subjects where verbal facility is needed, and in those subjects or that aspect of his subjects only, then his requirements become much clearer than if we enumerated an unsystematized and seemingly incongruous catalogue of subjects (handwork, drawing, writing, arithmetic, etc.) in which he does well, and another detailed list in which he does badly (reading, spelling, composition), without being able to specify any characters in common. And generally, if such far-reaching qualities as good memory, good motor co-ordination, weak visualization, poor auditory imagery, can be established as 'factors' having a fairly wide range, each entering not into a single subject or test, but into a group of tests or subjects, the grouping thus revealed will manifestly render both diagnosis and treatment more speedy and more effective.

The practical value of descriptions in terms of what is virtually a general factor (as in the awards, for example, of the junior county scholarship examinations, or in the 'I.Q.' or 'mental age' derived from tests of intelligence) is already well appreciated: the practical value of these less general descriptions, in terms of what are called 'group-factors,' is not so widely recognized. Yet, to my mind, it furnishes one of the most pressing reasons for research along factorial lines.[1] What is true in educational diagnosis is equally true in vocational diagnosis: unless group-factors can be established, vocational guidance, except as regards the general level of the career advised, becomes almost impracticable. Similarly, if such 'temperamental' factors as general emotionality, introversion, extraversion, depression, and the like can be successfully established, the diagnosis and treatment of the neurotic, the psychotic, and the delinquent will be greatly simplified.

[1] This was, indeed, the pressing reason which led me to an early search for factors other than general intelligence on the cognitive side. A severe adherence to the 'two-factor theory' (which recognizes no broad factors except that of general ability or *g*, eked out by innumerable specifics) would make educational guidance difficult and vocational guidance exceedingly complex. The clinical examination of backward pupils, and the beneficial effects of specialized treatment in cases of 'special disability,' strongly suggested the existence of group-factors; and statistically controlled investigations seemed requisite to verify these first-hand impressions.

(*b*) But the 'practical necessity' (to use Thomson's phrase) is not the only motive. The more obvious reason for expressing our description in terms of factors is the theoretical requirement that dictates the mode of description almost everywhere pursued in science—the desire, namely, for the increased logical clearness and cogency attained by using a *few*, *permanent*, and *independent* terms of reference, instead of a large, indefinite number of casual and semi-dependent concepts, changing from one problem to another. Whether there are four chemical elements, or ninety, or only two, the description of material substances is greatly simplified if the chemist can analyse them all into a limited number of independent and unaltering constituents. So in psychology. The traits we can observe, the tests we can apply, are numberless; those actually selected vary greatly from this investigator to that; and yet, as the correlations show, they are to a large extent functionally dependent on each other: in short, they overlap. To convert these correlated measurements of arbitrary and changing traits into terms of uncorrelated components, appearing and reappearing in successive investigations, would not only effect an enormous economy of thought, but (what is far more important) would greatly enhance the precision and the validity of our logical arguments.

But, before going farther, let us note that the comparison of psychological analysis with chemical analysis, though often invoked by earlier writers, has sometimes proved unintentionally misleading. When we analyse table salt into sodium, chlorine, and a residuum of impurities, we effect an actual physical separation; and we consequently infer that the component atoms or elements are as concrete as the particles of salt. With some such analogy in his mind, the student of factor-analysis in psychology is tempted to reify the factors named, and to visualize a logical analysis as a physical separation, tacitly assuming that, if distinct abilities are ever to be discovered, they will be concrete and separable 'organs,' like the heart or the lungs, and that the 'mental mechanisms' which form them will be localized in separate brain-centres or cortical areas.

In psychology, however—and, personally, I should add in

chemistry as well—what we are really analysing are not substances, but properties: and by properties (as will be seen later on) I understand not attributes inherent in substances, but simply relations manifested under certain constant or standard conditions. Thus, the 'verbal factor' is not necessarily to be identified with a 'verbal centre' in the brain or a 'verbal faculty' in the mind, any more than 'acidity' is to be identified with a special chemical substance or a special chemical force. It would be better, therefore, to seek some other analogy.

Nor is it difficult to find one, for every science exploits the same procedure. To take an example familiar to every schoolboy, we may compare the advantages of using independent factors in psychology with those of using latitude and longitude in geography, where, instead of stating that such and such a place lies so many miles or kilometres in such a direction from the place at which the speaker happens at the moment to be standing, we say that it is so many degrees north or south of one arbitrary line, and so many degrees east or west of another arbitrary line drawn at right angles to the first. There is no visible 'line' to be crossed at the equator; there are no concrete 'poles' to distinguish the northerly or the southerly directions from all the rest. Yet the unreality of the lines and points that are marked down upon our maps does not destroy their utility. In the same way, to disprove the concrete existence of a psychological factor is not—as is so often supposed—to abolish its scientific value or validity.

Our factors, therefore, are to be thought of in the first instance as lines or terms of reference only, not as concrete psychological entities. In order to give an adequate description of persons we must first discover in what *independent* directions a person may vary, and, at the same time, so far as possible, choose the direction so that each may carry with it a maximum amount of *dependent* variation.

The early descriptions of plants in the ancient herbals simply seized on their most conspicuous, if superficial, modes of variation: the size of the specimen, and the colour and the shape of its flowers. Here are three directions, apparently uncorrelated, in which a plant can vary, the variation in one respect being independent of variation

in the other. We might, therefore, choose size, colour, and shape as our three main 'botanical factors.' But these factors do not bring with them numerous secondary variations of which they are the cause or the diagnostic clue: we can infer nothing from them about the structure of the stem, of the leaves, etc.: these would consequently have to be enumerated in separate detail. The modern botanist chooses other 'factors,' less obvious perhaps but more systematically connected with the whole nature of the plant: the presence or absence of seeds, of flowers, of one or of two or more seed-leaves, of united or disunited stamens and petals, and so forth. At each stage he seeks to adopt the one principle of variation which will contain within itself a maximum amount of descriptive implication, or, as the factorist would put it, will *account for a maximum amount of variance*. Clearly, the psychologist should begin by following the same procedure: he describes a child as of a 'non-verbal type,' not because he believes in a 'centre for words,' but (i) because such a description implies a number of other peculiarities in regard to a large group of school subjects, and (ii) because at the same time it does not needlessly and confusingly overlap with descriptions referring to arithmetical or to manual skill.

Let us observe, too, that our factors enable us, not only to describe persons, but also to describe traits. And, as we shall see in a moment, the logical *description* or 'definition' of traits is closely related to the factorial *classification* of traits. Having distinguished a factor of general emotionality from that of general intelligence, a factor of repression from its opposite, and a factor of pleasantness from its opposite, and so on, we can define any particular trait that we like to take as our *infima species*, in terms of the subaltern *genera* indicated by these successive dichotomous classifications: we can, for example, define fear as an unpleasant, repressive, emotional tendency aroused by such and such stimuli and leading to such and such actions.

At the same time, it will be noted, the factorist's descriptions remain more or less empirical, and need therefore to be elucidated by non-factorial research. The modern biologist does not describe his specimens in terms suggested solely by the principle of maximum variation. With increased physiological and evolutionary knowledge the dependent traits are seen to be functionally and not merely empirically dependent on the fundamental factors. Psychology, however, has as yet scarcely passed the Linnæan stage. We have long known, for instance, how to describe all visual sensations in terms of three primary factors only; yet everyone recognizes that such factorial work is a mere preliminary: much

histological, physiological, and biochemical research is needed to turn the empirical factorization into an intelligible factorization, and even (as with the duplicity theory) to correct the inferences from mere quantitative analysis. And if we know so little about the physiology of colour vision, we know still less about the physiology of intellectual aptitudes and temperamental traits.[1]

(II) *Prediction.*—So long as the correlations on which our descriptive terminology is based remain mere statistical statements of co-existence (or tendencies to co-existence), and are not yet fully explicable from a functional standpoint, any inference based upon factorial results must be subject to all the limitations of an empirical induction. In consequence, however much they may be manipulated or transformed in the course of factorial analysis, the set of measurements derived from the study of a single sample remains in itself nothing more than a description of the performances of the particular pupils tested with the particular tests employed. That holds true even if the correlations which lead to, or are deduced from, the hypothetical factors are made to bridge an interval of time, and take the form of a prediction.

Suppose, for example, we apply a series of educational tests to a hundred nine-year-olds, wait five, ten, or twenty years, and then correlate the results of our early tests with the children's later progress at school, or at the university, or in after-life: the correlations still remain descriptive of the group that has been followed up, and of that alone. They would enable us to reconstruct what we already know, namely, the subsequent achievements of these particular pupils; but in themselves they can tell us little or nothing about other groups or other individuals.

There is, however, a widespread notion that, even though the discovery of a correlation by itself may not justify general predictions, nevertheless the discovery of

[1] In passing, may I point out that factorial work was attempted in psychology long before factor-analysis as such was recognized as involving a special correlational technique. The topics instanced in the text provide familiar examples. Thus, the work on the colour-triangle was essentially a problem in factor-analysis; and the reader will be able to provide numerous other examples for himself (see below, p. 84).

the underlying factors will. Thus, a research student will compute the average or summed performances of a group of children, first in a composite series of tests, such as the Binet scale, and secondly, after an interval, in a later composite examination, such as that for junior county scholarships; he then proceeds to calculate the correlation between the two. If he has already had experience of such calculations, he will know that the correlation between marks and measurements like these is a highly unstable figure, fluctuating with the standard deviation of the group, and ranging far more widely than its sampling error would suggest. And he will be rather chary of treating it as a basis for his forecasts. Let him read, however, that such a correlation is due to a 'common factor,' and let him identify this 'common factor' with something nameable, such as *g* (thus thinking of it as '*general* ability' rather than '*average* ability') or as 'intelligence' (which tacitly suggests '*innate* ability'); and he at once feels that he has got down to something far more solid than a mere descriptive coefficient: he will assume that both his correlation and his initial test-results rest on a firm and permanent foundation, and that this foundation will remain to support him even when he turns to offer opinions about some other group.

It seems important, therefore, to emphasize two points: first, that unless, in labelling the factor or by some other means, additional premisses are surreptitiously introduced, the factor extracted from a single set of correlations can claim no deeper reality than can be claimed by the correlations themselves; and, secondly, that a single set of correlations in its turn can of itself rarely afford a valid basis for inductive inferences. These points will require elucidation in some detail. Let me take the second first.

(A) *The Requirements of Inductive Inference.*—The inexperienced beginner still commonly supposes that, if a correlation is 'statistically significant' as judged by its sampling error, then it can straightway be generalized, and taken as applying to other groups; and more than one well-known investigator, who would doubtless be fully alive to this fallacy in dealing with a single correlation, has

dropped into the same error when dealing with conclusions suggested by a tabulated pattern of correlations. Under what special conditions, then, if at all, are we justified in making such generalizations ?

Most psychologists who have used factor-analysis, at any rate in this country, have regarded it as providing a means of inferring, from experiments on typical groups, fundamental laws of cognition holding good of all mental processes wherever they are found. To derive laws from limited facts, to argue from the behaviour of the sample to the behaviour of the total population, is to argue from the particular to the general ; it is, in short, to reason by induction. And many of the criticisms laid at the door of the factorist prove, when closely considered, to be criticisms, not (as is generally supposed) of the mathematical method of analysis adopted, but rather of the logic, or lack of logic, which the factorial arguments display. What Fisher says of the mathematicians' views about the analysis of variance is equally true of psychologists' views about analysis by factors : " many, if pressed, would say it is not possible to argue rigorously from the particular to the general, that all such arguments must involve some sort of guesswork, which they might admit to be plausible guesswork, but the rationale of which they would be unwilling to discuss." [1] Yet, as he himself adds, though we may frankly acknowledge that all such inferences " must be attended with uncertainty, this is not the same as to admit that such inferences cannot be absolutely rigorous."

Accordingly, it seems worth while considering in some detail why the inferences and the predictions of the psychological factorist appear so lacking in rigour. I propose, therefore, to examine the chief postulates or principles which, implicitly or explicitly, have been relied upon by factorial writers to increase the credibility of their conclusions, and incidentally I shall endeavour to set down what, to my view at any rate, are the commonest logical defects of such work. As a rule, the supposed weakness of

[1] [109], p. 4. Thurstone, for example, declares : " It is in the nature of science that no scientific law can ever be proved to be right : it can only be shown to be plausible " ([84], p. 45).

the inductive 'guesswork' has been fortified by three different appeals—first, by emphasizing the simplicity of the conclusions drawn; secondly, by emphasizing the positive analogies within the phenomena compared; thirdly, by relying on *a priori* plausibility to supplement the *a posteriori* inferences.

(*a*) *The Appeal to Simplicity.—Natura est simplex*, said Newton; and factorists commonly begin by declaring that the aim of factor-analysis, as of every form of scientific analysis, is to discover the simplest possible formulation of the facts.[1] Conversely, when a simplified formulation has been successfully attained, its very simplicity is supposed to guarantee its truth—a guarantee which could never be claimed on inductive principles alone. Thus, the tables of correlations met with in psychology often show patterns

[1] The section 'On the Nature of Science' with which Thurstone opens his statement of the 'factor problem' ([84], chap. I, pp. 44 *et seq.*) suggests this standpoint. "To discover a scientific law," he says, "is merely to discover that a man-made scheme serves to unify, and thereby to simplify, comprehension of a certain class of natural phenomena." Similarly, Kelley, in introducing his 'new method of analysis,' defends it on the ground of 'simplicity,' and adds: "to create such simplicity is a basic purpose of factorization" ([84], p. 3). Analogous phrases could be quoted from Spearman, Thomson, Guilford, and many others who have discussed the aims of factor-analysis.

The same postulate is more particularly invoked where the mathematical analysis alone would not lead to a unique solution. A striking instance, as we shall see, is Thurstone's proposal to accept that particular mathematical solution which conforms to the requirements of 'simple structure.' This is in keeping with a well-known practice of physicists. Thus, Jeffreys, in analysing experimental data obtained to illustrate the quantitative laws of mechanics, observes that, as a matter of fact, the law or formula that every physicist would accept "is only one of an infinite number of laws that would fit the data equally well: its special quality, that distinguishes it from the other possible laws, is its *simplicity*" (*Scientific Inference*, p. 37: his italics.) A more general discussion of the problem from a logical standpoint is undertaken by Johnson in connexion with 'functional induction.' "The mathematician," he writes, would "point out that there are an infinity of different functions that would exactly fit any finite number of cases of covariation. . . . To escape this threatening annihilation of inductive inference, we may indicate two fundamental criteria . . .: first, reliance is placed upon the character of the formula itself, in particular on its comparative simplicity; second, its higher credibility depends upon its analogies with other well-established formulæ" (*Logic*, Pt. II, chap. xi, pp. 250–1).

of striking regularity or simplicity: the 'hierarchical' pattern, where every row of coefficients is proportional to every other, so that the whole set can be explained in terms of a single factor, is the best-known instance;[1] the 'bipolar' pattern, where the entire table can be arranged in four quadrants, two positive and two negative, is another case.[2] Now, the emergence of such simple patterns, so it is commonly argued, can hardly be ascribed to chance; it must therefore constitute a significant item of evidence in favour of some underlying factor.

No doubt, in many experimental results, the very simplicity of a pattern of figures is rightly held to be suggestive of something more than could be inferred from the figures taken alone or individually; and the simplicity of a formula, at any rate in the simpler sciences, is always deemed an added reason for its acceptance. Yet, without extraneous information, it is seldom possible to say with certainty what that something is: for, strangely enough, the simplicity of a pattern or a formula may imply either a very small number of large causes or (what is so often ignored) a very large number of small causes.[3]

[1] See Appendix I, Table I.

[2] A typical example is seen in conclusions drawn from the symmetrical pattern formed by residual correlations after a factor has been eliminated (see Appendix I, Table V). In a research quoted by Stephenson ([97], p. 360) a similar bipolar pattern was found in factorizing a set of cognitive tests, when the number of items correlated was only twelve. Not one of the coefficients was statistically significant as judged by the ordinary sampling error; nevertheless, Stephenson maintains that the mere presence of the pattern ('system 5,' as he calls it) is of itself convincing evidence for the existence of two antithetical or 'obverse' factors. Now, as I have tried to show elsewhere, the consistency conditions, which every table of intercorrelations is bound to fulfil, tend inevitably to introduce such bipolar patterns. The pattern itself, therefore, affords no evidence for any factor other than chance, since it constitutes the most likely result where chance alone is operative. The assumption that bipolar symmetry cannot be a haphazard product seems analogous to the contention of a naïve student who, after calculating an average, added up all his positive residuals and found the total to be $+635$, and then, adding up all the negative residuals, found that these came to exactly -635; his exclamation was that such a remarkable coincidence between the two figures could not possibly be produced by accidental deviations.

[3] The latter alternative arises in physics as well as in psychology, and seems

Accordingly, we must, I think, distinguish between simplicity as an *explanatory* principle and simplicity as an *inferential* principle. Where evidence is limited and the phenomena are more or less complex, inferences and predictions based on simple formulæ are likely to have a higher probability than those based on formulæ that are more elaborate, merely because they make fewer arbitrary assumptions. But this does not mean that simple explanations, in themselves and as such, necessarily have a higher probability. Indeed, if we apply this line of reasoning to psychology, our explanations and the remoter inferences we shall be tempted to draw from them will generally be farther from the truth instead of nearer to it. Thus I myself should argue that, if simplicity is a reason for the acceptance of an explanation in a simple science, simplicity is a reason against its acceptance in a science whose subject-matter is highly complex.

Let us glance for a moment at a particular problem to see how the point arises. It has often been said that a hundred factorial theories could be advanced that would fit the psychologist's correlation tables quite as well as Spearman's two-factor hypothesis. But that does not necessarily refute the hypothesis. A hundred theories could be devised to predict the apparent movements of the planets across the sky. But that is no reason for rejecting the hypotheses of Copernicus or Newton: their astronomical theories, we are told, are really accepted because in that particular field there is a high *a priori* probability in favour of a simple explanation rather than a complex. The question therefore is: are we still

curiously forgotten by those psychologists who would model psychological science on physical. "If," says Poincaré, "we study the progress of scientific inquiry, we see two opposite phenomena: sometimes there is a simplicity concealed beneath complex appearances (as in the Newtonian laws explaining the movements of the planets); sometimes, however, the simplicity is apparent only and conceals realities that are extremely complicated (as in the superficially simple law of Mariotte, describing the kinetic phenomena of gases)." Psychologists of the single-factor school who appeal to the analogy of the 'simpler' Copernican hypothesis overlook the analogy of the gases, where a simplicity resembling that of the single-factor theory is really the result of numerous small erratic movements. As Poincaré observes: "Here the simplicity is apparent only, the product of an average result, and the grossness of our observations alone prevents us from perceiving the real complexity" (*La Science et l'Hypothèse*, p. 175).

warranted in allotting a higher *a priori* probability to the simpler theory when we are dealing, not with the comparatively simple phenomena of astronomy, but with the far more complex phenomena of mental life ?

A living creature is not a simple homogeneous ball, and its relevant environment is as complicated as that of a planet is simple. To suppose that the laws governing the movements of a planet whirling in almost empty space can be expressed in three concise equations containing only one or two terms and only low powers is plausible enough ; but to expect that the laws governing the movements of a wasp as it buzzes round a room will be equally few and simple is the reverse of plausible. Again and again, the history of pseudo-scientific theories in psychology has shown that the commonest reason for accepting an erroneous explanation in the past has been the strong popular prejudice in favour of simple and single explanations where highly elaborate explanations would be far more appropriate. The very simplicity of the ' two-factor ' hypothesis has given it a widespread popularity among students and teachers, but at the same time has led—or ought to have led—the neurologist, the biochemist, and the geneticist to be highly sceptical of its finality. And much the same is true of most of the speculative simplifications introduced by this school of psychology or that.

This point of view gains some empirical confirmation from the results of practical work. The three-factor theory of intellectual abilities, which admits group-factors as well as the general and the specific, is on the surface less simple than the two-factor theory, which virtually excludes group-factors : yet in giving prognoses for the development of subnormal children at a clinic, the former appears to lead to far fewer errors, so far as can be judged from the after-histories of the individual cases. We are told that the factors deduced by the method of ' principal components ' are less in keeping with the ' law of parsimony ' than the set of factors conforming with the requirements of a ' simple structure ' ; and presumably the same objection would lie against the method of least squares (of which the method of principal components is a special form). Yet, on re-analysing old data by various alternative procedures, I find that predictions based on the method of least squares are nearly always the more accurate.[1] And, regarded as a basis for

[1] This result was obtained by J. F. Steele and Miss E. R. Woodhead in a recent unpublished research, where various modes of factor-analysis were tried with the same set of clinical data, and the inferences checked by clinical after-histories. As we apply it, the ' method of least squares ' or ' weighted

predictive formulæ rather than as explanatory hypotheses, I myself should contend that the factors derived by the former method are in some ways more simple rather than less (just as, to a scientific eye, Einstein's formulæ are simpler than Newton's); for the factors are independent, and the first two or three (which alone have a high statistical significance) will always account for, and predict, a greater amount of variance than any two or three factors that fit a 'simple structure.' [1]

In psychology the simplicity we have to look for is not an *a priori* simplicity, but an empirically ascertained simplicity. As in other sciences we design our inquiries so as to secure the nearest approach to isolated systems, that is, so as to deal with one problem at a time. But, although in the simpler sciences like astronomy, we may often assume that the group of observations we are analysing form an approximately isolated system, in psychology such an assumption is likely to be highly precarious, even when the most carefully planned precautions have been taken. To choose tests or traits almost at random, and note the simplicity of the resulting correlational pattern, can mean little or nothing, except that the choice was nearly random. To select tests or traits (or, it may be, persons) according to some definable principle, and then show that a simple formula will summarize the results, may mean something: it provides at least a presumption that we have perceived what was relevant and eliminated what was irrelevant. Whether this is really so, however, can hardly be decided from a single experiment alone. We may, of course, invent methods of factorial research that will always yield a factor-pattern showing some degree of 'hierarchical' formation or (if we prefer) what is sometimes called 'simple structure.' But again the results will mean little or nothing: using the former, we could almost always demonstrate that a general factor

summation' inserts in each table its own appropriate diagonal elements, whereas the method of principal components treats them as invariably equal to unity (see Appendix I). Except with a few small tables, however, the difference in the diagonal elements has little effect on the results: as the figures plainly showed, it was rather the subsequent process of rotation that seemed to reduce the accuracy of the deductions.

[1] Cf. the typical result obtained below, Appendix I, Table XI.

exists; using the latter, we could almost always demonstrate, even with the same set of data, that it does not exist. The economy animating such inventions seems to be an economy in the number of samples quite as much as an economy in the number of factors: the factorist has in effect asked himself, what method can I apply so as to reach a unique conclusion on the basis of one correlation table only? But that, as it seems to me, is to misconceive the requirements of inductive arguments.

(*b*) *The Appeal to the Positive Analogy.*—There is thus a second way in which reliance on simplicity seems often to mislead the factorist. In physical inquiries, when an investigator obtains a simple pattern of figures from a single set of observations (as he does, for example, in observing the acceleration of a dropped weight), the simplicity of the resulting formula is held to absolve him from the need for repeating his experiment over and over again before he begins to generalize. Similarly in psychological inquiries, when a simple pattern has been obtained, the factorist is usually ready to generalize from one or two correlation tables only.

Once again he forgets the wide differences between the two sciences: the physicist[1] can generally assume in advance that a simple analysis will fit his simple material and that in his experiments he is really varying one factor only at a time; usually, indeed, he has deliberately selected a factor whose relation to the effects he is studying is likely to be expressible by some simple mathematical function.

[1] This is rather the physicist of the elementary textbook (who is held up in other sciences as a model investigator), not the actual research worker. It is quite exceptional for an experimenter in the research laboratory to vary only one factor at a time or to imagine that he is doing so. All that the textbook means is that, for purposes of simple exposition, his logical argument can be stated as if he had actually varied the factors one at a time, and that a rough demonstration can be arranged in the schoolroom to illustrate the main principle. Observe that the schoolboy, knowing nothing of the irrelevant factors, accepts a single instance as conclusive, and the proof as he understands it involves an 'intuitive' rather than a 'statistical' induction: but even in physical research experiments are multiplied more to reduce errors of measurement than because a mere increase of number leads to greater generality (cf. Johnson, *Logic*, Pt. II, chap. x, p. 216).

To that extent he may be fully justified in accepting an intelligible conclusion on the strength of a single experiment. But in psychology how can we ever be sure that our experiment is based on a complete analysis of the situation, and that all other conceivable influences have been successfully excluded save the one with which our hypothesis is concerned? As a rule, we do not even know what those other influences are. Our only hope of eliminating them is to repeat the whole experiment again and again with what we take to be irrelevant influences differing as widely as possible in each successive trial: in technical language, our argument must be based on the 'negative analogy' as well as on the 'positive.'

My point can be made clearer if I apply it to an actual example. For this purpose I shall choose a research where the fallacies stand out in flagrant relief. The same fallacies, I believe, often lurk in the publications of more authoritative writers, but naturally they there occur in more subtle and less obvious forms: to take a more competent piece of work would therefore obscure and complicate my illustration rather than render it plainer.

A thesis that I have before me seems especially appropriate, because the investigator (an able teacher and student who has graduated in psychology and logic) expressly appeals to logical principles. His main problem is formulated as follows: "What is the chief cause of intellectual progress? Do the educational achievements of our pupils depend upon the narrow instruction given in the ordinary classroom, or on the development of some wider psychological function? I shall answer," he continues, "that all teaching should be based upon exactly the opposite procedure from that hitherto adopted: instead of giving the pupil connexions between ideas to learn by heart, we should require the pupil to discover those connexions. This revolutionary principle follows immediately from a theory which can be experimentally proved, but which will at first no doubt sound scarcely credible, the theory, namely, that all educational progress depends on a single process, the process of educing relations between actions, concrete things, or abstract ideas. To demonstrate this theory it will obviously

be necessary to apply separate tests of educational achievements and of relational activities to one and the same group of pupils, and show that those who do well in the latter usually do well in the former."

Accordingly, he applies seven 'educational' and twelve 'psychological' group tests to fifty-three boys, aged 10–11 years, in two classes at an elementary school. The correlations, he finds, are "all significant, and with few exceptions would fit a hierarchy within three times the probable error." Since that margin allows a deviation of $\pm$ ·20, the alleged fit is an exceedingly loose one, and indeed is not explicitly invoked to support the conclusions drawn. His main argument may be given in his own words:

"The correlation between every pair of tests turns out to be both positive and significant. Hence every pair has a common factor. . . . The factor common to all the tests,[1] educational as well as psychological, is the perception of relations between abstract or concrete ideas. Thus our theory is verified; and we conclude that this relational ability is the general and essential cause, not only of success in the cognitive tests of relation-finding, but also of all intellectual progress at school. . . . It may therefore be maintained that every child will make the greatest amount of progress if, from the earliest ages on, he is taught, not by mere mechanical memorization as at present, but by a logical procedure which should enable him to understand the reasons for what he is taught, in a word, by teaching him to educe relations instead of to reproduce associations."

There is here a rather startling transition. We suddenly pass from the abstract notion of a common factor, invoked to account for a statistical correspondence, to the concrete notion of a causal factor, which can be exploited as an

[1] It is here tacitly assumed that the factor common to one pair is identical with the factor common to the other pairs. In his introductory chapter the writer states that "it is unscientific to postulate a number of factors when one factor will suffice." Presumably, therefore, his identification rests rather on the principle of simplicity or parsimony than on any theory of the cause of a hierarchical order among correlation coefficients: indeed, since he declares that his table contains "possible specific (i.e. group) factors," he cannot claim it as conclusively hierarchical.

effective means towards certain desirable ends. How is this transition bridged ? The logical link is indicated in the writer's introductory account of his statistical procedure. Here he describes correlation as " a method of measuring the agreement between two variables " ; and then, like other writers before him, quotes (or, rather, significantly misquotes) Mill's canon : " The method of agreement assures us that, if two or more instances of a phenomenon have one factor in common, that common factor is the cause of the given phenomenon." According to the axiom of universal causation, therefore, we tacitly *assume* that a common factor is a common cause. I shall later consider how far factorial arguments necessitate some such causal postulate. Meanwhile, let us note how the method of agreement is used in the problem before us. It furnishes what the investigator calls a ' decisive verification.' " Our generalization," he writes, " appears confirmed beyond dispute, when we examine the correlations of each test and of each school subject with the common factor, i.e. as they are called, their loadings or saturations, which indicate the extent of their agreement with that common factor. All the saturations are positive, and, with one exception, all are significant. Thus high achievement at school and high ability in the factor everywhere coincide. . . . We are consequently able to explain all intellectual progress at school as the effect of one simple noegenetic law." [1]

But is this the only conceivable explanation ? I observe that the instructions to each of the writer's group tests were given in abstract verbal terms ; and, from my own experience of group-testing, I suspect that the understanding of the instructions presented far greater difficulties to the young examinees than the solution of the actual test-problems. Accordingly I might argue with equal plausibility : " the understanding of abstract verbal instructions is a factor common to all the tests ; it is therefore the general cause of success in both the psychological and the educational problems."

[1] The writer has previously argued that " the third noegenetic law " (the eduction of correlates) " really depends on the second " (the eduction of relations) and that " the first . . . is not a noegenetic law at all."

The headmaster gives my argument a different turn. He writes : " those boys who did best were the younger and more conscientious workers " ; and for him, it appears, the " essential cause of success " is rather a quality of character than of intelligence—a general factor of conscientiousness. The pupils' own class teacher offers yet another explanation : the student has declared that there is a factor common to the educational tests and the psychological tests, and infers that it is the ability measured by the latter that constitutes the common cause ; the teacher, accepting the same premiss, puts cause and effect the other way round : it is, he urges, not the psychological ability that has produced the educational skill ; it is the skill measured by the educational tests, i.e. the skill imparted by his own teaching, that has enabled the pupils to solve the psychological problems.

If, however, instead of dividing the tests into educational and psychological, we classify them according to the concrete nature of the various problems, we can discern quite different common factors. Some of the tests, both the psychological and the educational, involve material that is primarily visual ; others depends largely on memory ; others again require the child to formulate an answer in words of his own. This suggests a threefold group-factor pattern conforming with the requirements of ' simple structure ' ; and, on actual trial, this pattern yields a far closer fit to the observed correlations than the hypothetical figures deduced from the writer's own saturation coefficients. On this basis, therefore, taking each group of tests in turn, we could argue with equal justice that, not the general factor common to all the tests, but the particular factor common to each limited group was the " essential cause of success " within that limited group.

The main conclusion of the thesis, therefore, is by no means " confirmed beyond dispute." There are half a dozen other explanations which would account for the results just as plausibly. Where precisely, then, has the fallacy crept in ? At bottom it arises from two time-honoured errors as to the nature of inductive reasoning, errors which crop up again and again in factorial arguments : first, induction is treated as a procedure for reaching certain

rather than probable generalizations; secondly, it is treated as a positive rather than as a negative procedure—as being based on the agreement[1] of positive instances instead of on elimination by means of negative instances. These misconceptions are so common in factorial work that I may be pardoned for exposing them in some detail: for unless we can fit factor-analysis in its true logical setting, we shall, I am convinced, utterly misconstrue its nature, and be continually led astray.

Where, as in psychology, the issue is somewhat involved, it is, I think, a helpful practice to encourage the young research student to outline for himself the successive steps in his argument so that the formal aspect of his reasoning shall be obvious at a glance. Any illicit transition will then leap to the eye. Here, assuming for the moment that the generalization to be proved is a certain and not a probable proposition, the essential premisses and conclusions may be set out schematically as follows. The writer begins with a hypothetical syllogism.

"If relation-finding is the cause of educational progress in these children, then their performances in the two types of test should agree;

"The correlations show that they do agree;

"Therefore, relation-finding is the cause of their progress."

The fallacy is plain. It could be succinctly if somewhat pedantically pinned down with the labels of scholastic logic: the writer argues in the *modus ponendo ponens* (and commits the fallacy of affirming the consequent), whereas he ought

[1] The 'method of agreement,' to which the writer (like so many other factorists) more than once appeals, is somewhat deceptively named. It assumes, not merely (as correlationists who cite Mill appear to suppose) that "the instances have one factor in common," but that (in Mill's own words) they have "*only* one circumstance in common" (*System of Logic*, Bk. III, chap. viii, p. 428). Its value therefore depends more on disagreement than on agreement. As Mill himself admits, the method "is an inferior resource" (p. 433), and the required assumption cannot be made unless we show we have "excluded all other causes." In other words, it is only valid when the 'relevant known positive analogy' (to use the terminology adopted below) is equal to the 'total analogy'; and this can hardly ever be the case in psychology, though it may seem to be the case in elementary mechanics.

to argue in the *modus tollendo ponens* (and proceed by disproving all the alternatives except one).

To this hypothetical argument is added a categorical syllogism in the third figure,[1] which is equally invalid as it stands :

" The 53 boys, aged 10–11, whom I have chosen for my experiment, owe their progress to relation-finding ;

" The 53 boys, aged 10–11, whom I have chosen for my experiment, are school children ;

" Therefore, all school children owe their progress to relation-finding."

Again the reasoning is plainly fallacious : it involves, like all problematic inductions when forced into this shape, an illicit process of the minor.

Both these fallacies are constantly committed in theses on factor-analysis. It will therefore be instructive to ask what premisses would render the two arguments valid ; for that will at once indicate what assumptions the writers are unconsciously introducing.

(1) As it stands, the hypothetical syllogism can only be validated by adding the premiss : ' this hypothesis (relation-finding as a cause of progress) is the *only* hypothesis which could account for the consequences specified (agreement in test-performances).' Usually the beginner is himself unable to think of any alternative hypothesis, and so silently concludes that his *is* the only hypothesis. Here lies the advantage of the experienced supervisor, who can nearly always suggest a long list of rival explanations. In psychology, to assume that your own hypothesis is the *only* possible hypothesis is far too sweeping. But in certain cases it may be quite legitimate to claim that it has *a higher antecedent probability* than the rest.

This limited claim will not permit us to generalize with certitude. If, however, we convert the hypothetical syllogism into an inference by probabilities, no formal

[1] Proving, as Aristotle would have said, ' the major of the middle by means of the minor ' (*Anal. Prior.*, *β*, xxiv, 68b, pp. 15–29, where ' minor,' of course, means not the subject of the conclusion, but the term of minor or minimum generality). Aldrich's pupils would have called it substituting ' Darapta ' for ' Darapti.'

fallacy will be perpetrated. Our student's contention would then reduce itself to this : " the *a posteriori* verification of a consequence deducible from my hypothesis (or ' theory,' as he calls it) appreciably increases the probability of a hypothesis that was already probable *a priori*." Note that in such an argument the final probability must depend on three things[1] : (i) the relatively high *a priori* probability of the hypothesis to be proved ; (ii) the number of independent conclusions that can be deduced from it in advance and verified by experiment ; (iii) the relatively low *a priori* probability of those conclusions. The last point is constantly overlooked or misunderstood. To argue, as our investigator does, that the *hypothesis* is improbable *a priori* reduces the probability of the final conclusion. To argue that the deducible *consequences* were improbable *a priori* would increase its probability. We tend to believe a speculative theory, not because it is surprising in itself, but because it explains, or enables us to predict, facts that would otherwise surprise us.

(2) The categorical syllogism could be validated if we could add the premiss : " all school children owe their progress to the same cause." Once again this is the kind of sweeping assumption that most naive thinkers make, until its obvious inaccuracy is pointed out. Later on, however, we shall see reason to suppose that the assumption contains a larger element of truth than the stickler for formalities usually realizes. It implies a belief in homogeneous populations—natural kinds, natural types, and the like. Such a belief, no doubt, is untenable in its primitive form. Nevertheless, some such postulate is essential to all attempts to generalize from experience.[2] And once again we can in

[1] The ' criteria of problematic inference ' are fully discussed by Keynes, Broad, and Johnson in the volumes cited below.

[2] This requirement is seen most clearly in an important type of inductive argument which is rarely considered, viz. generalization from a single specimen. The physicist will argue : " All gold has the same atomic weight ; this specimen has an atomic weight of 197 : therefore all gold has that weight." And, as I have already remarked, if he goes on to repeat his test with other specimens, it is rather to eliminate experimental errors than to extend the enumeration on which his induction is based. *Ab uno disce omnes.* But why does this apply to the weight of gold, but not to the colour of swans or

certain cases substitute a modified statement which will facilitate an argument as to probabilities, though it will never lead to certitude, namely, that the groups of children and tests examined form *representative or random samples* of the total populations of persons and tests.

Extenuating claims of this kind are often explicitly put forward by the more argumentative writers in the introductory sections to their theses, and at times are even discussed at greater length than the experimental data themselves. Why, then, do the less cautious writers commit fallacies so transparent as the above? And why do the more cautious insist so strongly on these speculative additions to their proofs? Usually, I think, because both of them, like so many who work with factor-analysis, are anxious to reach a 'unique solution.' They want everything to be settled once for all by a single analysis of a single set of data collected in a single research. If, by the aid of implicit or explicit assumptions, a 'unique solution' can be directly attained, the investigator is relieved of two troublesome duties—that of repeating his inquiry again and again under varying conditions and that of estimating the final probability of the net result.

In my opinion, however, a 'unique solution' is an ideal to be approached slowly and from different sides by progressive delimitation: it is not a simple concrete fact to be discovered in a single step, if only we invoke the necessary postulate and choose the right procedure. This being so, we may lay down the following precept, which most experi-

crows? Why should we be chary of inferring the weight or the intelligence of all 10-year-old boys from the measurement of one specimen alone when one specimen of gold suffices? To find a satisfactory answer for the factorist would lead us from logical into metaphysical considerations, and must therefore be postponed for the moment. But, although the answer may appreciably modify the popular notions, it must inevitably tend to support rather than refute the existence of relatively homogeneous populations or 'types.' Evidently, when we are dealing with the co-existence of mental attributes as distinct from physical, it is not so easy to convince ourselves that our specimens belong to such relatively homogeneous 'populations' or 'types.' But the very fact that we venture to make such inferences at all implies that something like distinguishable types (or tendencies making for such types) must be assumed. See below, pp. 70, 224.

enced investigators would perhaps think too obvious to formulate, but which is nevertheless constantly infringed : instead of attempting an argument which seeks to prove a universal certainty, be content to build up an argument in terms of probability.

In general, I suggest, the reasoning will take a disjunctive form, for example :

" The pattern of positive correlations disclosed by the performances of these children is due either (i) to chance, or (ii) to the fact that relation-finding is a factor common to the educational as well as to the psychological tests (a hypothesis for which there is a high antecedent probability), or (iii) to some other common factor or series of common factors (hypotheses for which the antecedent probabilities are not so high).

" Now, (i) the tests of significance render it highly improbable that the correlations (or at any rate the major portion of the variance and covariance those correlations imply) can be due entirely to chance ;

" (ii) I have selected such different types of the same tests and such different specimens of the total school population, that it is highly improbable that any other factor besides relation-finding can be common to all the test-performances " (or, "so far as any other factor was conceivable in the raw data, its effects have been eliminated by partial correlation, by mutual cancellation, or by some other equivalent factorial method) ;

" Therefore, (iii) in the instances examined, the correlations between the two types of test (or at any rate such and such a proportion of the variance) must probably be due to the only factor remaining, namely, relation-finding. Calculation shows that the probability is of such and such a magnitude."

So far, the conclusion is still restricted to the cases actually examined. The investigator must accordingly proceed to show that the correlation-pattern, or the factor-variance on which his argument turns, is sufficiently stable through a wide variety of samples, and that the examined samples may be reasonably accepted as fair specimens of the total population, as regards both relevant and irrelevant

characteristics. With a certain degree of probability, which will increase not only with the number, but also with the variety of the samples, he will then be justified in generalizing from the instances examined to all instances, examined and unexamined alike.

In such an argument, it will be seen, the cogency of the final conclusion does not depend, as is popularly supposed, on merely increasing the amount of agreement in the relevant factors—the 'positive analogy,'[1] as it has been called. It depends far more on increasing what has been called the 'negative analogy,'[1] that is to say, on diminishing the points of agreement[2] that are to be ignored in the conclusion, and so increasing the amount of difference in all irrelevant factors. This may mean formally disproving the chief rival explanations in special sections of the research; or it may mean planning the main research so that the rival factors have no room to operate. In either case—*maior est vis instantiae negativae.*

I do not suggest that this full procedure is always indispensable. Different generalizations have different antecedent probabilities, and so require different ranges of favourable *a posteriori* evidence to attain an acceptable degree of final probability. As we shall show in a moment, some appeal to the *a priori* probabilities is inevitable. Here, as we have already noted, our investigator, by his initial claim that his hypothesis is 'revolutionary,' and at first sight 'scarcely credible,' has really made a convincing

[1] These terms are introduced by J. M. Keynes [43], pp. 223 *et seq.* The principles proposed would be clearer if, in addition to distinguishing, as his terminology does, between (A) the positive and (B) the negative, and between (1) the total and (2) the known analogies, we also adopted technical names for (*a*) the relevant and (*b*) the irrelevant analogies; and then, within the relevant, known, positive analogy, distinguished between (i) the implying or diagnostic analogy (Mill's 'cause') and (ii) the implied or inferable analogy (Mill's 'phenomenon' regarded as an 'effect').

[2] Points of agreement, it should be added, common not only to the entire group, but also to sub-groups. The plausibility of the criticisms advanced by Thomson and Thurstone against the advocates of a single general factor largely depends on the fact that investigators, who may seem to have eliminated all but one factor common to the entire group, do not exclude the influence of a mixed set of (sub-) group-factors.

proof harder, not easier: like every young student, he yearns to show that he is proving something unexpected, something that had hitherto seemed improbable *a priori*—at least to all except himself. The paradoxical character of his conclusion, however, on which he has laid so much stress, springs simply from its sweeping form: even Mill could have warned him that "the precariousness of the method of simple enumeration is in an inverse ratio to the largeness of the generalization" (*loc. cit.*, Bk. III, chap. xxi, § 3). Had he narrowed his 'inferable positive analogy' to something that had a fairly high *a priori* probability (e.g. explicitly limited his conclusion to children of a certain age, certain intelligence, certain fundamental acquirements, and to lessons of a certain type, instead of generalizing about all children and all school subjects), his inductive proof could have proceeded quite plausibly on a much narrower basis than I have proposed.[1]

But in any case, at some stage of the work, an elimination of competing hypotheses is essential. The difficulty is to be sure that our enumeration of the possible competitors is exhaustive. "Rejection or exclusion," as Bacon observes, "is quickly said; but the way to come at it is intricate and winding."[2] There would seem to be at least three ways which, separately or simultaneously employed, may in some measure increase its efficacy. All of them depend on much the same principle—namely, on so planning the experimental and statistical procedure that the unknown influences may be legitimately treated as 'chance factors.'

First, by employing an appropriate method of multiple factor-analysis, we can resolve the given table into a series of

[1] We need not, I think, altogether accept the arguments advanced by the Oxford logicians against the traditional notion of induction, viz. that it is an *entirely* indirect and negative procedure, that its sole principle "in all its forms is elimination"—the "exclusion of all alternatives but one" (Joseph, *loc. cit.*, p. 430; Cook Wilson, *Statement and Inference*, vol. II, p. 595). Their criticisms, I take it, are valid only against that kind of empirical induction which, like Mill's, aimed at universal certainties, instead of being content with merely probable conclusions, i.e. with verifying hypotheses which themselves already possess a reasonably high *a priori* certainty.

[2] *Novum Organon*, Bk. II, Aphorism xvi (beginning: "The first task of induction is rejection or exclusion . . .").

independent[1] factors, extracted in order of their contribution to the total variance. This may enable us to deduct those factors that are not relevant to our main conclusion. Thus, if we are seeking to establish the importance of a general factor for intelligence, we can partial out the influence of the group-factor for verbality from the verbal as contrasted with the non-verbal tests; if we are seeking to establish the presence of a factor for verbality, we can partial out the possible influence of the general factor: having done so, we can then proceed to disprove the 'null hypothesis,' and so show that a significant amount of residual variance can only be accounted for by the particular factor whose presence we desire to establish.

But, owing to the size of the probable errors, it will often be impossible to isolate in a single table more than two or three identifiable factors. In such cases we can fall back upon a second device. We may endeavour to carry out the elimination experimentally instead of statistically. By carefully selecting the tests or the persons we may at the very outset succeed in excluding those irrelevant factors that happen to be most easily identifiable.

Thirdly, the effect of minor and lesser known factors may be to some extent neutralized by systematic randomization.[2] This is an artifice which would seem peculiarly appropriate for investigations in the psychological and social sciences. It has been strangely neglected hitherto. Later, I shall give illustrations of its use in what are essentially factorial problems.

Each of these three principles requires that more syste-

[1] Arguments of the type I have outlined above become exceedingly complex if correlated or 'oblique' factors are employed (see pp. 263 f.). In such a case the issue would apparently turn on obtaining measures of stability for what Lexis terms 'dependent' or 'organic' series (*gebundene Reihen*; cf. 'Uber die Theorie der Stabilität statischer Reihen,' *Abhand. z. Moral Statistik*, 1903, pp. 170, etc.)—a problem which he considered all but insoluble for practical purposes. Although tests for statistical stability, much better than those proposed by Lexis, are now available, his arguments still seem to me worth reading, and to have been much neglected in this country.

[2] The modern methods of experimental design for the 'analysis of variance' with multiple criteria indicate how this may be effected (cf. Snedecor, *Statistical Methods*, p. 210; Fisher, [109], pp. 18 *et seq.*).

matic attention should be paid to the appropriate selection of tests and persons than is usually the case. The first principle requires us to select samples of tests and persons that shall be relatively homogeneous in the relevant characteristics; the second, to select samples homogeneous in the irrelevant traits, introducing a relevant discontinuity into the sub-groups; the third, to select samples appropriately heterogeneous in the irrelevant characteristics—a point that is far more frequently disregarded.

If the hypothesis to be tested is sufficiently definite, and if the alternative factors are comparatively few, such precautions should certainly lead to a fairly simple pattern both in the correlation table and in the factorial matrix. But even so, neither the precautions nor the ensuing simplicity will, as a rule, enable us to transcend the fact that our own table is merely one specimen taken from the enormous number of analogous tables that presumably await investigation. In general, therefore, to justify a factorial prediction, or any other inductive generalization from the figures obtained by factor-analysis, the first prerequisite must be to base conclusions, not on a single sample or a single set, but on a series of such sets and samples.

It is for this reason that I have elsewhere proposed criteria which may serve to test the stability of factors from one investigation to another. Of these the simplest is the 'symmetry criterion.' If R_1 and R_2 are two correlation or covariance matrices dependent upon the same dominant factors, then $R_1R_2 \doteqdot R_2R_1$, i.e. the product of the two matrices should be approximately symmetrical.[1]

At the same time, let me add that there are grounds (which again cannot here be set out in detail) for doubting whether the coefficient of correlation is after all the best measure on which our inductive predictions are to be based. Strong reasons can be adduced, to a large extent following from the arguments just given, for preferring covariance to correlation, wherever covariance is legitimately calculable, and for basing predictions and statements of probability upon the regression coefficients rather than upon the coefficients of correlation themselves. A correlation coefficient is descriptive solely of the set of figures on which it is based: it cannot profess to

[1] A worked example is given in Table IV [128], p. 68. Other instances are given in the earlier theses by Williams and Davies (cf. [119] and [130]).

measure a physical or objective phenomenon, as a regression coefficient or a covariance may under certain conditions claim to do.[1]

(*c*) *The Appeal to* a priori *Postulates.*—But, however many tables we collect, we still cannot legitimately extrapolate the results, unless some further assumption is made about the total population or universe which our samples are presumed to represent. This holds of all empirical prediction. And here once again we observe how the peculiar fallacies that invalidate so many factorial generalizations are, in fact, but special instances of the difficulties that surround every attempt at reasoning by induction. No inductive inference can be justified on formal grounds alone : certain material postulates, generally obscure and in most cases unexpressed, are essential to carry logical conviction. As we have already seen, any effort to reach probabilities by inverse reasoning implies antecedent or *a priori* probabilities, as well as the explicit or *a posteriori* probabilities supplied by the empirical research ; and, however much the mode of argument is recast, it seems wholly impossible to escape such initial assumptions.

In factorial work they usually take the form of certain general notions about the structure of the mind or the physiological working of the nervous system. Assumptions of this kind are avowedly introduced alike by Thomson and by Spearman. It is true that, when made explicit, the enthusiastic advocate of factorial statistics will often reject these non-statistical postulates as irrelevant to the procedure itself : Stephenson, for example, has more than once protested against " attempts to drag physiology into the picture." Other critics declare that the very search for factors common to groups of traits or tests is equivalent to invoking the discredited doctrine of faculties common to various mental processes : while others again protest that the mere notion that the mind can be dissected by a quan-

[1] I have ventured to criticize the exclusive reliance placed by psychologists on the method of correlation in one or two earlier papers (cf. [93], p. 247, [121], p. 170 f.). " Some people have been misled into the belief that correlation is the key to all the secrets of nature. In reality, its utility as a statistical method is narrowly limited ; furthermore, it is one of the most difficult of statistics to explain " (Snedecor, *Statistical Methods*, 1937, p. 128).

titative analysis into additive elements assumes an atomistic structure quite incompatible with modern conceptions of the wholeness of personality. Nevertheless, however much the critics may object to this assumption or that, without some set of antecedent postulates, tacit if not overt, no valid generalization can ever be reached.

Hence, in all factorial researches, it is essential that the unproved premisses, which lurk in the background of almost every investigator's inferences, should be brought out into the open, and frankly recognized for what they are.[1] Each author should scrutinize his arguments to see where he is going beyond reasoning of a mere formal or statistical type, and then, so far as he can, adduce explicit support for any additional assumptions that he may be making *a priori*.

Often these presuppositions are unconsciously embodied in the very way in which the data are selected. If, for example, our evidence for a general intellective factor is derived from correlations between tests relating chiefly to sensory discrimination, the conclusion may easily be drawn that general sensory discrimination and the general intellective factor are identical. If our tests are largely verbal or largely scholastic, we may discover, or seem to discover, that verbal tests or vocabulary tests make excellent tests of intelligence, or that intelligence usually goes with high educational attainments. Those who rely mainly on motor, practical, spatial, or other non-verbal tests, will reach the opposite conclusion.

Here, then, we note once again how vital it is, if factor-analysis is to be employed, to consider the design of the experiments by which the initial data are to be obtained, as well as the accuracy of the mathematical methods by which the data are to be analysed. Again and again, in studies of the 'general factor,' the research worker simply takes whatever mental tests have a high reliability and can be conveniently applied to an accessible group of examinees; he hopes to demonstrate his factors, in any set of observations,

[1] My own suggestions will be given, in general form, on a later page (pp. 222 f.). The more concrete physiological assumptions that seem to me to fit in with factorial work I have already outlined elsewhere (*Brit. J. Educ. Psych.*, IX, p. 192).

even when collected with little or no intelligible plan, if only he can apply some fool-proof criterion. On the contrary, as I tried to point out in an early paper, it is requisite in all such researches to see that the *tests or traits*, as well as the persons tested or observed, constitute a fair and systematically selected sample: according to the nature of the investigation, it was argued, they should either be "chosen so as to represent, so far as possible, all the typical aspects or levels of the mind"; or else (this holds most frequently of the persons, but may also hold of the traits) they should be expressly selected so as to form what is called a 'random' sample.[1] The two principles are really the same: for 'random' selection does not mean blind or careless selection, but "selection according to some precept or method which ensures that the mode of choice shall be irrelevant to the probability of the generalization to be established." And "our knowledge of this irrelevance is prior to the empirical establishment of the generalization," and therefore *a priori*, as the phrase has been used here (cf. [43], pp. 41 f., 281 f.).[2]

(B) *The Weakness of the Factorial Link.*—My second point, it may be remembered, was that the introduction of factors, deduced from correlations, does not of itself (as is

[1] In my first investigation on 'Experimental Tests of General Intelligence' I sought to lay down the former principle as essential to researches based on ordinary experimental tests ([16], p. 98): the latter principle is often more appropriate in researches based on observations or impressions, since these may be more numerous than tests.

[2] The special problems of sampling tests or traits are not unlike those that have arisen more recently in connexion with sampling persons or populations. Most factorists appear to assume that in both cases the method of random sampling is the only available or legitimate procedure; and by random sampling they understand the method of 'simple sampling' as described by Yule and others ([110], p. 350] rather than the method of systematic randomization as described by Fisher ([109], p. 20 f.). This view is taken, for example, by Stephenson in a recent paper ([136], p. 20). As I have indicated elsewhere, in studying the existence of mental or social 'types' or tendencies among the general public, some method of representative or stratified sampling has often to be substituted for the simple effort to sample the whole population at random (e.g. in surveys of school populations). Similarly, in selecting tests or traits, the method of simple random sampling must in general give way to some more elaborate method of systematic sampling by strata, levels, or other representative scheme, in which, no doubt, randomiza-

so often supposed) increase the precision and the probability of the predictions ultimately reached : rather it diminishes them. Here we meet in a technical form another common fallacy of popular science. In psychology we are most familiar with it in discussions on formal training : the nineteenth-century schoolmaster would only believe that the teaching of science could help his pupils to think scientifically in after-life when he had first persuaded himself that a scientific training at school would strengthen some enduring common factor, such as the ' faculty of reasoning.' Similarly, in more practical problems : the inexpert predicter always feels greater confidence if he can base his forecasts on some generalized concept. He prefers to reason from simple, universal rules, even if he somewhat inconsistently admits that his rules, " like all rules, have their exceptions."

But his faith in the superior validity of such inferences is sadly misplaced. As has so often been pointed out : " It is better to argue immediately from the given particular instances to the new instance than to argue by way of a major premiss : the conclusion is only probable in either case ; but acquires a higher probability by the former method than by the latter." [1] Provided the evidential data are the same, an ' eduction ' can always be drawn with higher

tion will play an essential part. I need not enlarge on the special problems here. The statistical issues have been discussed by A. N. Kaer, *Bull. de l'Inst. Intern. de Statistique*, IX, p. 176 f., and A. L. Bowley and A. Jensen, *ibid.*, XXII, pp. 355 f. ; cf. also J. Neyman, *J. Roy. Stat. Soc.*, XCVII, pp. 558 f.

[1] B. Russell, ' On the Notion of Cause,' *Proc. Arist. Soc.*, N.S. XIII, p. 197 ; cf. Johnson, *Logic*, Pt. III, chap. iv, p. 44. It is instructive to note that, in their controversy over the way inductive arguments are to be justified, Mill and Whately are both agreed on the erroneous doctrine which I have criticized in the text. (Their common error, I take it, is due to their ignoring the conditions of *probable* reasoning : they ignored them because, like everyone else at the time, they were ignorant of them.) Mill expressly states that, even in inductive arguments, " no conclusion is proved, for which there cannot be found a true major premiss." Thus, if we wish to infer by induction that *any* given person has this or that common attribute, we require (so he maintains), as " a necessary condition of the validity of the argument," the immediate major premiss " Whatever is true of John, Peter, etc., is true of all mankind " (*System of Logic*, Bk. III, chap. iii, pp. 343–4). This is still the implicit view of most inductive scientists, including nearly all psychological factorists.

probability than an ' induction ' or a ' deduction ' resting on an ' induction.'

The Relative Probabilities of Inductions and Eductions.—To the student this point will doubtless be clearer if he compares (*a*) the inductive-deductive procedure with (*b*) the eductive at three stages of increasing difficulty: (1) for verbal arguments of the traditional ' class type,' based on complete generalizations; (2) for statistical arguments of the ' attribute type,' based on proportions or probabilities; and finally (3) for statistical arguments of the ' variable type ' based on regressions.

(1) In their simplest forms the two modes of reasoning would follow some such schemes as these:

(*a*) *Inductive-Deductive.*—The samples examined show that " (i) All children who succeed in this test are intelligent. (ii) All children who are intelligent will gain scholarships." But " (iii) Individual *i* has succeeded in this test. Therefore, (iv) being intelligent, he will gain a scholarship."

(*b*) *Eductive.*—The sample examined shows that " (i) the children who do well in this test will gain a scholarship." But " (ii) *i* has succeeded in this test. Therefore, (iii) he will gain a scholarship."

Is it not obvious that an appeal to premiss *b* (i) affords a safer ground for the common conclusion than a twofold appeal to premisses *a* (i) and *a* (ii)? What woman, instead of trying on a new hat at first hand, would be content to send her dress to the milliner, and argue, " my dress matches my complexion; the hat matches my dress; therefore the hat must match my complexion "? Even the carpenter who carries a foot-rule prefers to make a direct comparison, instead of trusting to Euclid's first axiom.

(2) But in an empirical science our initial observations can hardly ever be comprised in complete and sweeping generalizations. Hence, instead of using categorical arguments of the traditional type, we must fall back on statistical arguments in terms of frequencies or probabilities. At this level, our premisses will take the form: (i) Judging by our sample, " Such and such a proportion of the children who succeed in the test will prove to be intelligent "; from which we deduce, (ii) " if *i* has succeeded, the proba-

bility that he is intelligent will be $p/h = \cdot 90$ (say) " ; from other data we infer that (iii) " if any particular child is intelligent, the probability of his obtaining a scholarship will be $q/ph = \cdot 80$ (say)." And so, by invoking the familiar multiplicative theorem,[1] we finally conclude : (iv) " the probability that *i* will be intelligent and so gain a scholarship is $p.q/h = p/h.q/ph$, i.e. $\cdot 90 \times \cdot 80 = \cdot 72$."

But once again is it not obvious, without formal proof, that a safer and quite possibly a higher figure could be attained by directly ascertaining what proportion among those who have succeeded in the test actually win scholarships? Now, however, we may note that there will be two cases : the relatively trivial case in which *i* is among those who have already been followed up, and the case of practical importance in which he is a member not of the sample actually observed, but only of the population sampled. It is the latter that provides the interesting problems for the logicians who study the nature of probable inference and for the statisticians who devise methods of estimation. As we shall see later on, in this latter case, the probability we ultimately assign must largely depend upon the particular *a priori* postulates we tacitly or explicitly invoke. This is a further complication to be borne in mind ; but it is simpler to disregard it for the moment.[2]

[1] The product theorem for probabilities is usually stated as an axiom or definition (Johnson, *loc. cit.*, p. 181 ; Keynes, *loc. cit.*, p. 135, cf. p. 148). It is proved as a theorem in Coolidge, *Probability*, p. 18 ; and the familiar proof is given a novel and suggestive turn, in keeping with the analogies indicated above, in Jeffreys, *Scientific Inference*, p. 17. The notation used in the text is that introduced by McColl and Keynes [43] and now fairly widely adopted : *p*, *q*, *h*, denote propositions ; *q/ph* denotes " the probability that *q* will be true on the assumption that *p* and *h* are both true." For the relation between statistical eduction and deduction from statistical generalizations (' inductions ') cf. Broad, *Proc. Arist. Soc.*, XXVIII, pp. 2 *et seq.*

[2] In the latter case the plausibility of predictions based on the common factor arises from the fact that most factorists identify the common or general factor with an ' ability ' which they assume to be ' innate ' and therefore constant. There is, of course, some justification for the latter assumptions. Nevertheless, with that interpretation we should not really be dealing with precisely the same mathematical ' factor ' at the two different moments of time : for the latter is itself only imperfectly correlated with ' innate intelligence.' In any case, such identifications overlook the important points that a person's factorial composition changes as he grows older and that the

(3) In factor-analysis, however, we deal not with attributes defining discrete classes, but with variables differing in degree. This brings us to arguments of a third type. What we have to infer is not the probability that *i* will be intelligent or win a scholarship; that is still too crude a formulation: what we require is an estimation of the most probable amount by which *i* will deviate above or below the average, first, as regards intelligence, and, secondly, in the marks at the scholarship examination.

By the eductive method we should calculate at a single step the correlation between the intelligence test (which we may call test 1) and the scholarship examination (which we may call test 2)—or rather the 'regression' of the latter on the former. If the marks for both are in standard measure, we have

$$m_{2i} = r_{21} \cdot m_{1i}$$

where m_{1i} is *i*'s mark in test 1, m_{2i} is the estimate of his mark in test 2, and r_{21} the coefficient of correlation. Here we may assume that, if *i* is not in the batch tested and followed up, his most probable mark in test 2 will be the average of those examinees in the tested sample who obtained the same mark as he in test 1.

By the inductive-deductive method we should proceed as follows. For simplicity we may continue to suppose that only one test has been used to measure intelligence, and that only one common factor, namely, *g*, is affecting both the measurements in the intelligence test and the marks in the scholarship examination.[1] Having first

factorial composition of the same set of tests is not only different for different individuals, but different at different ages.

[1] The student, I hope, will not assume that, if we had a *number* of tests incorporated into the intelligence and scholarship examinations respectively (instead of only one in each), that would of itself improve the *relative* trustworthiness of inferences based on the factor *g*; or that, if we also had several factors instead of one, that would necessarily give the indirect predictions a higher value: for, with matrix notation, all the arguments in the text can be generalized for as many variables in each set as we please. I may add that, if we want to find the best possible predictive coefficients enabling us to infer from one multiple test to another multiple test, we ought strictly to employ an entirely different procedure, which I have called bi-multiple correlation; but that procedure would again short-circuit the factorial deduction.

ascertained by factor-analysis the hypothetical correlations of g with m_1 and m_2, we should estimate i's most probable marks in two steps by regression equations as follows :

$$\begin{aligned} g_i &= r_{1g} \cdot m_{1i} ; \\ m_{2i} &= r_{2g} \cdot g_i \\ &= r_{2g}\, r_{1g} \cdot m_{1i}. \end{aligned}$$

The result is reversible ; for, by analogous regression equations, we could estimate i's probable measurement in the first test from his mark in the second, viz.

$$m_{1i} = r_{1g}\, r_{2g} \cdot m_{2i}.$$

The correlation coefficient being the geometric mean of the two regressions, we obtain

$$r_{12} = r_{1g}\, r_{2g}$$

or (with the alternative notation) $= f_{11} \cdot f_{21}$.

The Product Theorem.—This last equation yields what I have termed the ' product theorem.' It may be regarded as the analogue of the multiplicative theorem in simple probability.[1] It is, as we shall discover in a moment, the

[1] Both have close analogies to the logical multiplication of classes, relations, and propositions. Indeed, I am tempted to say that the product theorems mentioned in the text are but particular cases of the more general product theorem that forms the basis of all deductive logic (if R_{ab} and R_{bc} denote given relations between a and b and between b and c respectively, then the relation between a and c will be defined by $R_{ac} = R_{ab} \times R_{bc}$; if A and B be two classes, defined by $\phi(x)$ and $\psi(x)$ respectively, then their common part C will be defined by $\phi(x) \times \psi(x)$—where in either case the multiplication symbol is defined in a more general way than is usual in finite arithmetical multiplication, but will include this as a particular case). Since I shall presently argue that the mathematical reasoning of the factorist, like all mathematical reasoning, is but a special example of formal reasoning generally, whether quantitative or non-quantitative, it may be worth while to exhibit the analogies more explicitly at this stage.

(1) Let us use the proper fraction $\frac{M}{S}$ to symbolize such propositions as " Some (or all) children who succeed in this test (S) are also intelligent (M)." We may then write the syllogism set out in 1 (a) above as follows :

$$\frac{\text{Number of children who are also intelligent}}{\text{Number of children who succeed in the test}} \times \frac{\text{Number of children who also gain scholarships}}{\text{Number of children who are intelligent}} = \frac{\text{Number of children who also gain scholarships}}{\text{Number of children who succeed in the test}}.$$

If (as is assumed in paragraph 1 in the text) the first premiss is universal, i.e. if *all* the children who succeed in the test are also intelligent, the value of

central theorem in all factorial work. I have stated it as follows: "If test 1 is correlated with the only common factor g to the extent of r_{1g} and test 2 is correlated with the same factor to the extent of r_{2g}, then test 1 and test 2 will be correlated amongst themselves to the extent of the product of those two correlations, namely, $r_{12} = r_{1g} \cdot r_{2g}$:" [1] more briefly, the correlation between two variables is the product of their correlations with the only common factor.

When we apply this theorem for purposes of prediction and the like, it is important to remember that we are dealing only with an *estimation* of r_{12} and the corresponding regression coefficient. No doubt, in theory we can show

the first fraction will be unity. Similarly if the second premiss is universal. It will then follow that the third fraction must also be unity, i.e. that *all* the children who succeed in the test will also gain scholarships, as the syllogism concludes.

(2) If (as we assume in paragraph 2) only *some* of the children who succeed in the test prove to be really intelligent, then we can insert the actual number, e.g. the average as deduced from our sample; and the first fraction will be less than unity. Similarly for the second premiss. Taking the probability to be the ratio of the two frequencies (or, more accurately, the limit approached by this ratio as the number in the denominator is increased indefinitely) we arrive at the multiplicative theorem stated in para. 2, p. 47. If the argument is kept in terms of frequencies, it is equivalent to that used in the association of attributes. Thus, employing Yule's notation, let A = number who succeed in the intelligence test, B = number who gain scholarships, C = number who are intelligent, AC = number who are intelligent among those who succeed in the test, and so on. Then, if intelligence is the only common factor, there will be no partial association between A and B within class C; and $\delta_{AB.C} = 0$. Accordingly, by Yule's formula, $\frac{(ABC)}{(A)} = \frac{(AC)}{(A)} \cdot \frac{(BC)}{(C)}$ (cf. Yule [25], p. 49; [110], pp. 56–7). In passing, it should be noted that the 'criterion for independence' within class C is really a criterion for showing that the factor responsible for the classification into C and not-C is the only common factor: we shall recur to this below (p. 147).

(3) Finally, with a slightly different line of reasoning, for "average number of (children who are also intelligent," etc.) we may substitute "average deviation of (the same children in intelligence)," with similar substitutions in the other fractions. We then reach the last form of the product theorem as stated in paragraph 3.

[1] *The Measurement of Mental Capacities*, pp. 11–12: cf. also *Brit. J. Psych.*, 1909, III, pp. 159–60; *L.C.C. Report*, 1917, p. 53, equation (ii). The application of the theorem is fully illustrated in each of these publications.

that, if we restrict our notion to the simplest kind of error, then, for the trivial case above mentioned, the amount of error involved in the indirect calculation is the same as that involved in the direct : i.e. we have imposed on ourselves the extra labour of a roundabout computation with no improvement in the result. But in practice the actual errors involved at the two stages will nearly always prove to be cumulative. Occasionally, no doubt, certain *a priori* postulates may be invoked which suggest that the error of the indirect method may be the smaller ; and, what is still more important, it may occasionally be more economical in applied psychology to fall back on the indirect method, because a little extra time spent over calculations may save a vast amount of time spent over experimental research. And this, as I shall show in a moment, is the chief justification for basing predictions on a factorial procedure. Otherwise, it is safe to say that reliance on an intermediary factor will never diminish, but may often increase, the errors of our prediction. Wherever we can, it is best to by-pass the factors.

Incidentally, let us note that, in demonstrating or using the product theorem itself, it is not necessary to assume (as both its employers and their critics generally do) that the common factor is actually a common cause. Indeed, one of the great dangers in attributing causal properties to such factors is that the assumption may encourage us to trust them too much. The correlations which we multiply and obtain by multiplication are best thought of as merely indicating probabilities for the purpose of an empirical forecast or estimate : and the multiplication itself plays the same part as in other arguments based on probabilities. 'Abilities' themselves cannot really be multiplied like forces.

Nevertheless, for the sake of a simplified exposition, it may be convenient to regard a correlation as a ratio stating what proportion of causal elements are common to the two correlated variables. The analogy between the product theorem and the multiplicative axiom is then still more obvious. To take an illustration I have used in an earlier *Report*, let us suppose we know the true order of merit (g) for a given set of examination scripts ; and let us suppose these scripts are marked by two examiners, A and B, who combine the true order with a certain amount of error, weighting it, let us say, in the proportions $c/t_a = 1/2$ and

$c/t_b = 1/3$. We can then imitate A's results by averaging the true order with a second order uncorrelated with it [1]; and we can imitate B's results by averaging the true order with two other orders uncorrelated with it and with each other. According to the simplest interpretation of correlation, if t_a is the total number of independent elements determining a variable a, and c the number of elements common both to it and to another variable g consisting wholly of those common elements, then it can easily be shown [2] that

$$r_{ag} = \sqrt{\frac{c}{t_a}}; \text{ similarly, } r_{bg} = \sqrt{\frac{c}{t_b}}; \text{ and}$$

$$r_{ab} = \frac{c}{\sqrt{t_a\, t_b}}: \text{ that is, } r_{ab} = \sqrt{\frac{c}{t_a}} \cdot \sqrt{\frac{c}{t_b}} = r_{ag} \cdot r_{bg}$$

[1] A worked illustration will be found in my *Note on Correlation as Applied to Mental Testing*, Board of Education Report [48], pp. 188–9.

[2] Cf. Bowley, *loc. cit.*, 1904, p. 355. These 'ratio-formulæ' are based on Bravais's method of deducing the correlation rather than Pearson's ('Analyse mathématique sur les probabilités de situation d'un point,' *Acad. Sci. : mém. sav.*, ii^{ème} sér., t.IX, 1846, pp. 255 f., eq. 28). Pearson's deduction of the product-moment formula proceeds by taking the product-sums of the empirical measurements (obtaining $r = mm'$); Bravais's original deduction proceeded by taking the product-sums of the independent elements of which those measurements were supposed to be composed (obtaining $r = ff'$). Bravais's mode of approaching the subject, though heavily criticized in the past (cf. [47], p. 152), is full of suggestive points for the factorist, which the other approach has led us to neglect.

The idea that correlation may be regarded as expressing the proportion of common elements has appealed especially to those interested in heredity. This was the origin of the celebrated experiments of Darbishire and Weldon on artificial correlation obtained by throwing dice [14]; and it still finds favour as a principle of teaching and interpretation (cf. Snedecor, *loc. cit. sup.*, pp. 128, 130: "Roughly this is the interpretation of the father-son correlation in stature, which is not far from ½: an average of some 50 per cent. of the genes are common to father and son."). Thomson adopts a similar standpoint in his 'sampling theory,' which he introduces with an ingenious extension of Weldon's experiments and formulæ ([132], p. 11). Here, however, it seems important to ask whether Thomson's elements ('neurone arcs') are as independent as the biologist's elements ('genes'): for the neurones belonging to the same individual have all developed from the same material. This, however, is a point we must take up later on.

Meanwhile, it may be noted that, if mental factors or 'abilities' are regarded as equivalent to forces, then the product-theorem is the equivalent of the familiar formula for the composition of two forces. This analogy, however, is also full of unsafe suggestions, unless its precise grounds are made clear (see below, p. 91). Indeed, I fancy it might be legitimately maintained

as may be roughly verified by direct calculation from the averaged orders.[1]

The fact that direct prediction is safer than indirect is quickly realized by any investigator who engages in educational and vocational guidance and is able to follow up his cases over an appreciable period of time. In earlier writings I have sought to stress this caution more than once. Nevertheless, these reservations have often been overlooked both by those who rely on psychometric methods and by those who criticize them. Accordingly, it may be advisable to consider in more concrete detail the two commonest cases in which attempts are made to base a practical prediction on

that the compounding of forces is itself a compounding of logical implications, not of actual causal entities. In psychology at all events, once the student has grasped the notion in the concrete, he should, I hold, look upon it as a principle of reasoning rather than as a principle of causation.

[1] The formal proof on which I have usually relied proceeds by writing $m_{ki} = \beta_{kg} \cdot g_i + e_{ki}$; the product theorem follows at once on correlating m_{ai} and m_{bi} ([93], p. 281, eq. xix). If we express the correlations as cosines and rewrite this initial equation $m_{ki} = \cos \theta_k \cdot g_i + \sin \theta_k \cdot s_{ki}$, the same argument brings out the analogy between what I may call 'factor-synthesis' and the composition of forces, and, conversely, the analogy between factor-analysis and the resolution of forces.

The shortest proof is that obtained from Yule's formula for partial correlation ([25], p. 238, eq. 12), by putting the residual correlation $r_{12 \cdot g} = 0$; but this formula really assumes the product-theorem to prove the residual correlation instead of *vice versa*, and the deduction does not make the theorem clearer to the student who has not followed Yule's somewhat lengthy demonstration. For the non-mathematical student the argument (with the illustrative exercise suggested in the text) seems to be the clearest. I relied upon it in my earliest paper because I then assumed that the correlation of a test with the common factor might be plausibly supposed to depend on and increase with its complexity. The instance there given is worth recalling because it illustrates the complexity theory of the general factor, which still seems to me to contain an important element of truth.

To quote my original example ([16], p. 160), suppose we have a series of sensori-motor functions, each of varying degrees of complication, yet all essentially manifestations of one common process, say motor co-ordination (X): for instance, we may imagine each to consist of a different number of elementary sensori-motor reactions, added or otherwise combined, so that the most complicated test, A, is equivalent to, say, a dozen determinations, and the most simple, B, to only three. Then, applying the ratio-formula, we can find a measure of the influence of X on A and B by computing its hypothetical correlation with each: and similarly, we can compute the

factorial generalizations, namely, what may be called educational and vocational prediction respectively. This will afford an opportunity of removing certain recurrent misunderstandings by the way, and of replying to typical criticisms.[1]

(*a*) *Prediction in Educational Psychology.*—With the aid of special tests and other devices, the educational psychologist seeks to discover, during their earliest years, first, those children who will never be able to profit by an ordinary elementary education, and, secondly, those children whose higher abilities deserve something more than an ordinary elementary education—the mentally defective and the future scholarship winners respectively. By way of illustration let us consider the second problem a little more closely. In theory the psychologist might begin by trying to discover what special mental qualifications are needed to pass the scholarship examination and to do satisfactory

probable intercorrelation directly. Thus, with the degrees of complexity just assumed, the correlation of A with the common factor would be twice as high as the correlation of B: consequently, A will correlate with all other motor processes, P and Q, twice as highly as B does; i.e. $r_{ap}/r_{bp} = r_{aq}/r_{bq}$: a result which can obviously be generalized for all ratios of the correlations with the common factor.

But this causal background, though convenient for a simplified exposition, is by no means necessary to a proof of the abstract theorem; and its somewhat speculative assumptions have been more than once called in question. Spearman, for example, in discussing my results, holds that the larger intercorrelations obtained with tests of higher and more complex mental processes are due simply to the fact that the "multitude of independent specifics must cancel one another, leaving the general factor more dominant" ([24], p. 69). However, the 'proportionality criterion,' as it may be called ('Burt's equation,' Spearman terms it), could itself be deduced by applying Spearman's own earlier formulæ—e.g. his formulæ for eliminating chance errors or for correcting 'distortion' ([12], pp. 90, 95), as indeed was pointed out at the time ([16], p. 159, footnote 3): he himself prefers to express the underlying principle in the form of a 'tetrad-difference equation' (see below, p. 149).

[1] The most recent criticisms are those of my friend and colleague in the L.C.C. Inspectorate, Dr. J. C. Hill ('A Criticism of Mental Testing,' *Brit. J. Med. Psychol.*, XVII, 1938, pp. 258–72, and previous communications). Dr. Hill complains that, in writing for the practical teacher, factorists like myself have laid far too much stress on alleged 'definite entities' such as (so-called) Intelligence: I ought not, he says, to condemn 19 out of 20 backward children, as being capable only of special education and of limited

work at a secondary school of the usual type; he might reduce these to a short list of non-overlapping 'fundamental traits' or 'factors'—memory, perhaps, and reasoning, verbal fluency, numerical accuracy, and no doubt certain moral and social characteristics that scarcely lend themselves to testing; he might then compile a number of tests to measure the more important intellectual factors in due proportion. Such a set of tests, or something very like it, is in fact comprised in the Binet-Simon scale and in many published booklets for the testing of 'educable capacity.' These compilations, however, were for the most part drawn up quite empirically. In actual practice I doubt whether any psychologist would first undertake a double factorization of the scholarship examination, on the one hand, and then of the Binet tests, on the other, and base his predictions on the agreement between the two sets of factors. He correlates performance at the Binet tests or at the group

industrial usefulness, solely on the ground of such indirect statistical inferences. Similarly, the writers of the two papers quoted below protest that the "interposed factors" introduce dubious and unnecessary complications into the educationist's deductions. Dr. Reed, for example, refers to my early *L.C.C. Memorandum* on Junior County Scholarships, and asks: "Does not Dr. Burt's whole argument break down if, with some psychologists, we doubt the very existence of special faculties or factors?" He advanced similar objections against the introduction of 'vocational guidance in the schools,' contending that 'vocational psychologists seem to be faculty psychologists without knowing it.' In reply I should like to draw my critic's attention to the 'Statistical Note' appended to the *Report* he quotes: he will there find a formal algebraic proof that "statistical inferences, mediated by hypothetical factors, cannot be more safe, and usually are less safe, than the determination of direct correlations"; for this reason, in the body of the report, "inferences with factors as middle terms" were employed only where no data existed for direct deduction. As for vocational guidance, I have always admitted that "vocational psychologists are perhaps too ready to assume what are variously called, according to the fashion of the moment, 'faculties,' 'specific capacities,' or 'factors,'" and have argued for a direct empirical procedure wherever possible rather than an indirect or analytic ('Principles of Vocational Guidance,' *Brit. J. Psychol.*, XIV, 1924, p. 351). Finally, may I say that Dr. Hill's article does not give a fair reflection of the factorist's views? Thus his main conclusion I myself should willingly accept. Indeed, the peroration of his article (last four lines of p. 270) is almost a verbatim reproduction, metaphor and all, of the concluding sentences of my own *Report* (reprinted in *The Subnormal Mind*, p. 134).

tests *directly* with performance in the scholarship examination, and relies on this direct correlation.

No doubt an application of factorial principles will aid in the construction and selection of tests; but it has often proved misleading. For example, a strong belief in the 'single-factor theory' (I myself should be inclined to say, a real misunderstanding of the 'single-factor theory') has led many of its supporters to aim at homogeneity rather than heterogeneity in their batteries of intelligence-tests. "It follows from Spearman's hierarchical principle," we are told, "that a reliable battery of tests for the general factor should have the highest possible correlations with each other . . . Thus the Binet scale stands condemned at a glance by the extremely miscellaneous nature of the tests that compose it: never having applied the factorial method, Binet evidently failed to make up his mind as to what precisely his tests were to measure." The point is put still more strongly by Cattell in his recent defence of 'the factorial methods of analysis of personality': "if, as most psychologists concede, the [Binet] test is not concerned with one ability but with a collection of abilities, the attachment of a single quantitative value to this hodge-podge is meaningless." [1]

These criticisms once again exhibit the fallacy of trusting solely to the 'positive analogy' to strengthen inductive inferences. The right guiding principle I have endeavoured to state more than once elsewhere. It is the logician's principle of increasing the 'negative analogy.' "Multiplying the number of different examiners is of greatest value when their correlation with the 'true mark' is high and their correlation with each other is low . . . The same holds true of the subjects tested: the best results are obtained by combining tests which correlate highly with the general ability to be measured, but attack it from independent or divergent angles." [2]

[1] *Character and Personality*, VI, 1937, p. 115: on p. 121 he argues in detail against 'the error of the view' that factors are mere 'middlemen between tests and criterion.' On the issue discussed in the following paragraph I find myself in close sympathy with many of Cattell's conclusions, but on the wider issue I agree still more closely with those of Vernon as expressed in the same symposium.

[2] *Marks of Examiners*, pp. 304, 310. *Mental and Scholastic Tests*, p. 207. The guiding principle cited above is implicitly followed by boards of examiners when they arrange that their several question papers, though bearing one and all upon the central subject, shall nevertheless each deal with a widely different aspect. This is commonly defended on the favourite ground of economy—to avoid superfluous overlapping: its real defence rests on the inductive principle described above (pp. 29 f.).

Thus, the heterogeneity of the tests in the Binet scale, so far from diminishing its value, actually increases it; and the popular batteries of written group tests, which are put forward as better substitutes, and in which each test claims high correlations with the rest, are, as a rule, far too homogeneous: they owe their high correlation quite as much to what, with a more varied assortment of test-material, would prove to be 'overlapping specifics' (ease in understanding the printed word, facility of verbal expression, and the like) as to the 'general factor' of intelligence.

In educational work the psychologist's predictions often cover a long range. In certifying the mentally defective we are by statute required to show that the defect has existed "from birth or from an early age"; by the age of 6 or 7 we endeavour to discover those who during their school career will never be able to profit by an ordinary elementary education; towards the age of 10 or 11 we seek to select others whose subsequent educational progress will justify the award of a scholarship to a secondary school. In all these instances of testing, examining, or diagnosing, we are trying to ascertain, not (as the detective does) who has committed a momentary act, but who possesses a given amount of some lasting capacity. Since these educational predictions cover most forms of intellectual work, the capacity we desire to estimate must be a 'general factor'; and since the predictions envisage the rest of the child's educational career and indirectly the rest of his working life, this factor must have stability or permanence. We thus need something like a first law of motion for mental activity: that is, we require to distinguish a persisting internal 'state of uniform motion,' on which our predictions shall be based, from unforeseen 'external forces,' which may subsequently come into action and partly obscure or disturb it. Ideally we should like to detect this permanent internal 'state' when the child first comes into our ken, i.e. 'at birth or at an early age.' Such a state, conferred at birth and lasting throughout life, is precisely what is meant by an innate capacity or propensity; and the external forces are represented by environmental influences, such as training at home, teaching at school, illness, accident, and the like.

Accordingly, from the very outset of my educational work it has seemed essential, not merely to show that a general factor underlies the cognitive group of mental activities, but also that this general factor (or some important component of it) is innate or permanent, in the sense just defined. In dealing with other tasks of psychological guidance—the treatment of delinquency or neurotic tendencies or the subsequent choice of employment—similar issues arise: we

want to know, not merely what factors are general, but how far those general factors are stable. In each case the problem is primarily one for correlation and statistical analysis. The precise methods that may be employed, and the provisional results so far secured, I have already discussed in early writings.[1] Here, therefore, I need do no more than emphasize this somewhat neglected aspect of factorial work.

Again, in examining individual children we often seek, not only to predict each child's general progress, but also to analyse out the special weaknesses of this child or of that. For such a purpose the heterogeneity of the Binet tests brings with it an added advantage: for they often throw an incidental light on these more specialized characteristics. As time goes on we shall come more and more to rely on the results of factorial work.[2] But in these cases the

[1] [16], p. 170 f., [22], p. 15 f., [23], p. 250 f. In the more popular formulation of my conclusions I freely availed myself of biological terms, and spoke of these tendencies as 'inborn,' as often 'hereditary,' and as identifiable on the conative side with 'instinctive' trends. This has led to occasional criticism, which is, I fancy, directed rather against the associations which those terms suggest, than against the ideas I intended to convey. I fully agree that the distinction between what is innate and what is acquired is an abstract and artificial distinction: so is the distinction between a projectile's 'state of uniform motion' and the 'external forces' that compel it to deviate from a straight line. I equally agree that the physiological and biological evidence for innate neural dispositions, and for genes governing those dispositions, is at present inconclusive: yet, so far as it goes, it lends some support to inferences from correlational work, and even throws some tentative light upon the nature of the more permanent factors. Nor does the weakness of the biological evidence destroy the correlational evidence, which is the chief thing on which the practical psychologist ought to rely. Only by correlational studies shall we be able to determine how far we can safely predict the future performances of a child at a subsequent age from his performances when he is first tested or, it may be, from the performances of his parents, his brothers and sisters, and his other relatives. Only by experimental checks—removing a child to a new environment, improving his physical health, altering the way he is taught—shall we be able to assure ourselves whether the factors we presume to be permanent—intelligence, for example, as tested by this or that set of tests—are really permanent and unchangeable (see *The Backward Child*, pp. 540–1).

[2] A misstatement in the interesting article by Cattell, just cited, needs correction, namely, that (except for a study of his own dealing with tests for the ages below eight) "the test-items (in the Binet scale) have never been validated by statistical treatment" (*loc. cit.*, p. 115). In the work that he quotes and criticizes (*Mental and Scholastic Tests*) a separate correlation coefficient (based on a simplified tetrachoric formula) indicating the 'validity' of every test-item is given (Table XXXI, p. 205). In an earlier 'Annual

practical problem is again rather different. It is not so much to predict the future as to decide on the immediate treatment—a matter to which I shall return in a moment; and I cannot help thinking that our critics have confused the logical requirements of two different issues.[1]

Report of the Psychologist to the London County Council' (1922) a more detailed factor-analysis of the results of the London Revision of the Binet-Simon tests was attempted by means of partial correlation and the 'group-factor method.' (The results are briefly indicated on pp. 184 and 195 of *Mental and Scholastic Tests*.) Besides the general factor, we found the same group-factors as were previously reported in the analysis of scholastic tests —viz., verbal, numerical, and manual, and, in addition, a memory-factor, a visual or spatial factor, a relational or reasoning factor, and a series of factors depending upon special knowledge acquired at home or at school, e.g. familiarity with money, etc. Several of my educational colleagues who have criticized the earlier report (including Dr. Hill) evidently assumed that the object of the factor-analysis was to justify the use of the scale as a test of intelligence; that, however, was dealt with on the basis of direct correlation in another section (pp. 199 *et seq.* and Table XXXI, as cited above). The purpose of the analysis into group-factors was rather to discover "how far the scale could be used for the incidental diagnosis of more specialized mental abilities." The results were by no means wholly favourable to the Binet scale, particularly in its original form. A similar analysis is being carried out jointly by Miss Simmins, Miss Davidson, and myself for the new Terman-Merrill revision; and this may provide us with an opportunity for replying more fully to our critics.

[1] This confusion seems discernible, not only in the articles by Hill and Cattell cited above, but also in two critical papers, which I have recently received, by Dr. R. F. Reed and by Profs. H. A. Reyburn and J. G. Taylor. Like Hill, Reed regards "the appeal to factors as inserting the weakest possible link into a chain that is none too strong." "All the educational predictions of the factorist," he argues, "are purely hypothetical . . .; he first assumes that the factor extracted from his tests is identical with what teachers call intelligence, and then he assumes that limitations in hereditary intelligence as thus tested will necessarily limit school progress, regardless of what the teacher and the school medical officer can do. Thus, Dr. Burt, having tested Arthur and found his mental ratio to be less than 85, assures the teacher that Arthur's case is hopeless and that Arthur is backward for life." But Reed, like Hill, entirely overlooks the fact that both the 'assumptions' that he attributes to the factorist have also been verified by direct correlation. Besides extracting a 'general factor' by analysis, I also correlated it, first with independent assessments of intelligence by competent teachers, and secondly (in showing that it was largely 'hereditary') with the intelligence of the children's parents; and finally, since I hold that factors as such need not come into the picture, I correlated the intelligence tests *directly* with the children's subsequent achievements. As for Arthur and the other children named in the chapter criticized by Reed, they were kept under observation

(*b*) *Prediction in Vocational Psychology*.—In vocational psychology, as in educational psychology, there seems also to be a confusion of two different problems. Here the first

for many years; and correlations between the performances of these and other children at their first testing and their progress later on are given in detail. Since the correlations are not perfect, it was never claimed that school progress is "*necessarily* limited" by the indications of the tests.

Reyburn and Taylor ('Factorial Analysis and School Subjects: A Criticism,' to appear in a forthcoming number of the *Brit. J. Educ. Psychol.*) deal more particularly with special abilities and disabilities. Their criticism is mainly directed against my article on 'The Relations of Educational Abilities' (*Brit. J. Educ. Psychol.*, IX, pp. 45–70). They overlook the fact that this paper was primarily concerned with a theoretical comparison of factorial methods, and that the concrete results were cited solely by way of confirmation of earlier statistical *and clinical* work (see opening paragraphs of that article). They argue that "the schoolmaster . . . would gladly learn of factors which would help him to estimate how the boy will do in some other examination, and more particularly to control them and alter their effects." They then treat these two problems—future prediction and immediate treatment—as on the same footing. As regards prediction they contend that "the interposed factors are an unnecessary complication: they are not recognized in the further situations to which they are a guide." With the first sentence I agree (with the reservations implied above in the text); with the second sentence I disagree: in the subsequent histories of my cases the factors *are* recognizable. As regards the treatment of individual children and the 'control and alteration' of their educational performances, the writers argue that the statistical methods I have used can have little or no value, because the resulting factors can be "psychologically and educationally significant only by accident." I should agree—*if* the correlated tests had been chosen without reference to educational problems, e.g. if (as in some researches which the writers have in mind) they were practically a haphazard collection. As it is, the tests were expressly selected to elicit those particular types of disability that had already been observed in clinical work. The writers finally conclude that, in view of the many alternative analyses that can be made of the same table (for they do not admit that the group-factor method and the general-factor method give equivalent results) "we doubt whether a satisfactory analysis of school subjects can be made with the available data." My reply would be that the "available data" do not consist solely of the table of correlations with which they are concerned. The 'available data' also include clinical studies of typical cases, experimental attempts to 'control and alter' educational disabilities by different modes of teaching and training, etc. etc., such as are described in my book. It is on the mutual illumination of both methods of approach that the practical educationist relies. As I have repeatedly insisted, tests, even if based on the most careful factorial work, "can still be but the beginning, never the end, of the examination of the individual child" (*Mental and Scholastic Tests*, p. xv; cf. *The Backward Child*, pp. 63 *et seq.*).

problem is the typical problem of vocational selection, the second the typical problem of vocational guidance.

Let us consider an actual example of vocational selection first. An investigator desires to measure the predictive value of a set of 'æstheto-kinetic' tests that he has devised for vocational selection. He may follow one of two procedures, which I have elsewhere called the 'analytic' and the 'empirical' respectively.[1] Adopting the 'analytic' procedure, he will, by special experiments or on the basis of general impression, attempt a factorization first of the performances involved in the vocational task, then of the performances involved in the tests, with a view to showing that the mental aptitudes required are essentially the same. Probably he will begin by applying factorial or other methods to analyse the measured skill of a group of trade workers into its component mental functions. He finds, let us suppose, that the chief functions are intelligence, sensory discrimination, and motor dexterity. Accordingly, he selects or constructs tests intended to measure these underlying abilities, applies them to a sample group of testees in his laboratory, and demonstrates by factor-analysis that the results of his tests depend essentially on the same three factors as the trade skill. From this he infers that the set of tests he has drawn up may be used to predict the vocational success of future workers in the trade concerned.

With the 'empirical' procedure he will apply his tests directly to the workers themselves, and deduce a regression

[1] I have ventured to abridge and quote passages from one or two early memoranda drawn up while I was head of the Vocational Department of the National Institute of Industrial Psychology. When that department was first founded, it was my duty to draft a note on general principles for the first investigations: and I am indebted to the Council of the Institute for permission to make use of reports and other memoranda written while I was working there. Like many other principles in vocational psychology (which, owing to the difficulty of obtaining older children for experimental research, can only be verified gradually and with difficulty) those described in the text were in the first instance largely deduced as an application to vocational problems of guiding ideas that had been found useful in previous educational work. Though some of the suggestions proved impracticable, the points here made have, I think, stood the test of experience. The original arguments will be found set out in fuller detail in the chapter and article cited below.

equation (or its equivalent) for estimating success at the work from the tests, without the mediation of any hypothetical factors. This, indeed, is the procedure that I myself have always advocated wherever conditions permit. Since estimation inevitably involves an error at every stage, the more stages we introduce into the total chain of estimations, the greater will be the cumulative error in the final prediction. Hence, as in educational selection, so in vocational selection, when dealing with any specific case, the empirically ascertained regressions will always be more trustworthy than indirectly reconstructed inferences based on hypothetical factors.

But once again the analytic or factorial approach is not without its value. One great advantage, as it seems to me, of analysing out the supposed factors and giving them names is that the isolation and the naming force us to see that our choice of tests covers a sufficiently wide range. An empirical selection is not always so mechanical as it professes to be: unconscious preferences tend often to narrow its scope. But this is simply to echo the maxim that I have so often emphasized in educational work—namely, that the psychological investigator with his tests and observations must be sure that he covers all the different aspects both of the child and of the child's task. Or, to reword it in statistical terms, he must see that the various qualities that he proposes to measure, although each of them is highly correlated with efficiency at the job, are not themselves correlated one with another (except, of course, so far as some small degree of correlation can scarcely be avoided).[1]

But the factorial approach has a further advantage in vocational work, which though present is not so striking in the field of education. It makes for economy of thought, and, what is still more important in practice, for economy of labour. To devise tests and to undertake researches for every conceivable vocation and for every conceivable group is not a feasible proposal. Hence the psychologist hopes that both the initial test-results and the ultimate vocational performances may always be reducible to terms of the same

[1] See 'Some Principles of Vocational Guidance,' *Brit. J. Psych.*, XIV, 1924, pp. 344 *et seq.* (Cf. also Thomson [132], pp. 114 *et seq.*)

limited number of factors.[1] He assumes, as a working hypothesis, that both occupations and individuals will prove to be roughly classifiable into ' natural groups ' or ' types,' and that the groupings to some extent will correspond.

This is still more obvious when we turn from the problem of ' selection ' to the problem of ' guidance.' For vocational guidance some form of multiple factorization seems almost essential. To offer guidance on any general scale we require in principle to relate all the aptitudes that may be displayed by any conceivable child to all the essential qualifications that may be required in any conceivable vocation. Such a comprehensive scheme of correspondences will be practicable only if both tested aptitudes and desirable qualifications can be reduced to terms of a few fundamental concepts. In terms of such concepts our regression equations (or their equivalent, a systematic case-study or ' psychogram ' for each child, a systematic job-analysis for each vocation) should be able to specify, on the one hand, the complex character of the individual, and, on the other hand, the complex requirements of the employment contemplated.

The initial difficulty and the solution proposed may be clearer if they are put in the following way. In an early study of the practicability of vocational guidance we found from the census that there were something like 700 different occupations into which boys and girls might ultimately be sent ((53), pp. 3 *et seq.*). We began, therefore, by classifying them according to the grade of intelligence each required ; and then proceeded to sub-classify them on the basis of special abilities, temperamental traits, and other broad requirements. For the sake of argument let us imagine that by proceeding in this way we could discover 6 independent factors or ' key-qualities ' that will account for the greater part of this variance ; and let us assume that it is sufficient to distinguish for each factor three grades only—average, above average, below average. This will yield $3^6 = 729$ pigeonholes into which all our cases may in theory be sorted. Hence, if each one of the combinations of grades corresponds with the main requirements in one of the specified vocations, the 700 vocations would be more than covered. Of

[1] Burt, *Vocational Diagnosis in Industry and at School*, *ap.* Muscio, *Lectures on Industrial Administration*, pp. 99 *et seq.*

course, I do not seriously contend that there is anything like a one-to-one correspondence. But it seems clear that, only by following some such general plan—the method of ' progressive delimitation,' as it was originally termed [1]—can the task be reduced to manageable dimensions.

But now, be it noticed, the function of the factors is not so much prediction as description : they aim at a first broad classification of children and of occupations according to what we have termed their vocational type. Moreover, any prediction or recommendation derived solely from factorial specifications would be so wide and mechanical that it would certainly need to be supplemented by a qualitative study of a more intensive kind. Hence, as we argued in our *Report*, once a tentative classification in terms of factors has been made, " the problem in its final stage should be turned into one of selection rather than of guidance " ; and here we must have recourse to direct estimation as before.

Some of the considerations I have brought forward in the last few pages may lead us to wonder whether after all the factors as such may not be otiose, and to inquire if eventually they may not altogether disappear from the psychological picture. Here perhaps it will be useful to distinguish between factors regarded as ' abilities ' and factors regarded as patterns of behaviour. To my mind, even when our arguments are couched in numerical form, the essential thing is not the supposed ' ability,' regarded as a self-subsistent quantity that can be measured like physical energy or power, but the pattern of correlations between the various hypothetical performances, or, if I may use more technical language, not the ' component ' as such, but the ' unit hierarchy ' that describes the component.[2] These patterns, these hierarchical matrices, are in

[1] [53], pp. 57 and 81–82. Thus, to take the factors in the order proposed in the *Report*, if a particular child is found to be in the middle grade for intelligence, in the upper grade for mechanical ability, and in the lowest grade for sociability, we should at once have reduced the number of vocations suitable for him from 700 to about 30 (chiefly skilled trades, not requiring co-operative work in the workshop). A further finer grading of intelligence, a further consideration of more specialized abilities and temperamental qualities, and a glance at reports on physical health, home background, previous training, and the like would probably diminish the number of appropriate openings to half a dozen or less (cf. *loc. cit.*, pp. 82–5 and Tables IV and XXXIV).

[2] Cf. ' The Unit Hierarchy and its Properties,' *Psychometrika*, III, pp. 151 *et seq.* It may be observed that, in discussing the *practical* applications of factor-analysis (e.g. to vocational guidance), Thomson himself has described ' factors ' as " unnecessary middlemen between the tests and the occupational

their turn nothing but algebraic devices for expressing complex qualitative wholes in quantitative form, in order to render our reasoning more rigorous.

We shall see in a moment that, in examining the assumptions made by certain logicians to account for inductive generalization in the physical sciences, Broad has already observed that it should be possible " to eliminate the hypothetical generating factors " (postulated for this purpose) " and to state the case wholly in terms of observable characteristics and their relations." His formal proof,[1] expressed in terms of attributes rather than of variables, could, I believe, be elaborated to prove that the same conclusion holds good in psychological factor-analysis. But what seems to have been overlooked by the critics of factors is this—a point strongly emphasized by Broad himself: " even if we confined our efforts to establishing *eductions*, and gave up efforts to establish *generalizations* inductively, we should still be pre-supposing the existence of universal laws."

(III) *Causal Explanation.*—The view which we have reached is very different from the view which most factorists hold—or at any rate from the view which most readers derive from the phraseology that factorists employ. Whereas we have concluded that mental factors have less objective importance than the actions from which they are inferred, most factorists apparently regard them as more real and more objective. Factors are accorded a superior predictive power, not only because they are tacitly assumed to possess a more concrete and more permanent nature than overt actions or behaviour, but also because they are held to be the true producers of the performances we observe and of the correlations between them. The latter are but

criterion " (*loc. cit. sup.*, pp. 114, 307, and refs.); his arguments tempt me to ask whether the ' factors ' could not also be suppressed in his *theoretical* deductions as well. It will be remembered that for Thomson the mental factors which our tests excite or sample are in the last analysis ' numerous small components.' But with the revised form of his ' sampling theory,' does not the postulate of organization into ' sub-pools ' become far more important than the postulate of ' atomic elements ' ? It would seem, indeed, that this suggestion is not far removed from Thomson's own view : for, as he himself observes, " the only reason for using the word ' elements ' is that it is difficult to speak of the different parts of the mind without assuming some ' items ' in terms of which to think."

[1] *Arist. Soc. Proc., loc. cit.*, pp. 38–9. See below, p. 224.

outward and visible effects: the factors (in Spearman's phrase) are the "hidden underlying causes."[1]

No doubt, this causal language, which we all to some extent favour, arises partly from the irrepressible disposition of the human mind to reify and even to personify whatever it can—to picture inferred reasons as realities and to endow those realities with an active force. But, in this particular reference, it is still further strengthened by the phraseology of most statisticians, who have nearly always discussed probability in terms of hypothetical 'causes.' From Laplace onwards, scientific writers on the theory of probability have regularly represented it as a procedure for 'inferring events from causes' or inversely for 'ascending from events to their causes.'

Thus, Bravais's original deduction of the product-moment function proceeded on this basis. Or again, to cite one of the earliest textbooks of statistics to introduce the subject of correlation, Bowley,[2] after expounding the general conception, formally deduces the equation $r = \frac{p}{p+q}$, and then concludes that "expressed in words the formula shows that the correlation coefficient tends to be the ratio of the number of *causes* common in the genesis of the two variables to the whole number of independent *causes* on which each depends." More recently, Fisher has given an identical formula to illustrate the 'analysis of variance' into 'portions contributed by the two *causes*'—the 'common' cause and the non-common: "in such cases," he writes, "the correlation merely measures the relative importance of two groups of factors *causing* variation" ([50], pp. 212, 210). Brown and Thomson have deduced

[1] Cf. *loc. cit. sup.*, p. 4. Actually Spearman's own system of factors has been criticized by Thomson because, though it "gives an admirable *description* of correlation data" (i.e. of "certain types of normal and abnormal persons"), "it does not give the *causes*" ([87], p. 64: italics as in the original). The latter, it is suggested, are supplied by the more numerous and more elementary components, which are "not new entities, but things we already know of," and in themselves far more simple "aspects of the causal background."

[2] *Elements of Statistics*, vol. II, p. 336 (my italics). See Keynes' trenchant criticism of this passage [43], p. 425.

a similar equation, and base their interpretation of the hierarchy upon it ; as a rule, however, they prefer to speak, not of causes, but of ' elements or factors ' ([39], p. 176).[1]

Accordingly, it is not surprising if psychologists should treat the ' common factors ' of one writer as synonymous with the ' common causes ' of another, or that they should go on to identify the ' factors ' by which the future actions of their testees can be predicted with concrete ' abilities ' or ' mental energies ' which are conceived as effective causal agencies determining such actions.[2] Some of them explicitly declare, as we have already seen, that the ' fundamental tendencies ' revealed by correlational analysis " account for, explain, and are the cause of, all human conduct."

Perhaps the most explicit expression of this view is contained in Cattell's article cited above. The alternative view—that ' different kinds of properties, belonging to human characters, can be distinguished, but not separated,' and that ' the attribute is not a part of the concrete individual, but only an aspect '—he emphatically rejects, classing it, with the ' tenets of much *Struktur* and *Gestalt* psychology,' as ' nothing less than a denial of the validity of scientific method.' Instead he argues that factors are ' psychological powers ' which can be measured, " like the power of a muscle," . . . " in interactionist terms of the real effects of those powers upon the physical and social environment," and are " eventually expressible in terms of energy transactions " (*loc. cit.*, pp. 127 *et seq.*).

The more recent statistical textbooks, it is true, abound in warnings to the student not to accept mere correlation

[1] Cf. Brown, *Mental Measurement*, 1911, p. 79 (where the equation is deduced as a generalization of Weldon's experiment illustrating correlation by combinations of dice). Cf. also p. 52 above.

[2] The transition is made possible by the ambiguities that lurk in the word ' cause.' The more cautious statistical writers expressly state the broad meaning they attach to the term. Thus Coolidge writes : " We shall mean by the *cause* of an event any *antecedent event whatever* " (*Probability*, 1925, p. 88, his italics). Even Coolidge, however, goes on to talk of causes as ' operative ' and as ' producing results ' (p. 88); and thus seems to forget his own definition and to attribute to his ' causes ' powers that his initial definition does not include.

as a sign of causation.[1] Yet psychologists still continue to assume that, even when the correlated traits may not themselves be causes, the factors deducible from such correlations must represent the causes that lie behind them. Once again, my own view is that what holds of correlation holds also of the underlying factor: its discovery might reinforce a pre-existing presumption in favour of causal connexion; but it cannot by itself create such a presumption. "The truth is, sensible investigators will only employ correlational methods to test or confirm conclusions at which they have arrived on other grounds."[2]

Thus, to borrow Thomson's language I should say, not merely that the 'causal entities' discovered by factor analysis *may* be "things we already know of in other connections" ([87], p. 89), but that, if they are entities at all, they *must* be "things we already know of in other connections," or at least things that we have antecedent reasons to postulate as probable or as convenient: their existence is in no way attested by the process of factorization. Hence the fundamental arguments supporting the existence or assumption of such entities fall outside the factorial proofs themelves, and so do not directly concern us at this stage. The most we can do will be, after reviewing the whole

[1] Kelley seems to express the true view most clearly when he describes correlation "as a measure of *mutual implication* and a measure *derived from* the regression coefficient" ([47], p. 189). Psychological writers, who at times appear to regard the statement of a correlation as the final goal of all statistical work, usually treat the regression coefficient as derived from a fundamental correlation instead of *vice versa*. Statisticians, on the other hand, would nowadays almost all subscribe to Fisher's view: correlation "will be found useful in the exploratory stages of an enquiry, . . . but, with controlled experimental conditions, it is seldom that it is desirable to express our conclusion in the form of a correlation coefficient." As I have indicated elsewhere, a correlation coefficient can easily be altered by selection, and is never the expression of a physical quantity as a regression coefficient may be. Generally, it seems best to regard both correlations and regressions as constants which, like the mean and the standard deviation, are merely descriptive of certain characteristics of the sample studied (see above, p. 15).

[2] J. M. Keynes, [43], p. 426. The whole of Keynes' discussion of correlation is worth re-reading for its bearing on the related problem of factor-analysis. Clear statements of the limitations of the correlation coefficient are to be found in Tippett, *Method of Statistics* (p. 126 *et seq.*), and Fisher ([50], p. 160 *et seq.*).

position, to consider whether the concrete correlational results (i) tend to verify, or (ii) are inconsistent with, or (iii) are completely independent of such causal theories. And this aspect of the problem will evidently have to be discussed in the light of metaphysical rather than of empirical considerations.[1]

In support of this contention it will perhaps be sufficient to appeal to the favourite logician of the factorist—J. S. Mill. Mill recognized, it may be remembered, 'uniformities of coexistence' as well as 'uniformities of causation.'[2] For Mill a causal uniformity is not a mere empirical relation, but a necessary relation. A cause, as distinct from a mere invariable coexistence, he defines as "the antecedent, or concurrence of antecedents, on which a phenomenon is invariably *and unconditionally* consequent"[3]: such uniformities, he seeks to show, can be proved by inductive arguments to be both certain and universally true. But with these universal causal certainties he explicitly contrasts what he describes as "approximate generalizations"—conclusions which are only probable, or only true (so far as we know) in some instances and not in all. Unless they in turn depend on causation, uniformities of coexistence are not necessary, but at most merely invariable. Hence they cannot as such be established with certainty by inductive arguments: for 'there is no general axiom standing in the same relation to the uniformities of coexistence as the law of causation does to those of succession.'[4] Consequently, they have the status of 'approximate generalizations' only. Now it is with these empirical 'uniformities of coexistence,' not with any alleged causal certainties, that factor-analysis, to my mind, is primarily concerned.

As the chief examples of uniformities of coexistence, Mill cites the regular conjunction of specific attributes or properties in what he calls 'Natural Kinds': e.g. (to take one of his instances) the coexistence of such characteristics as blackness of skin and woolliness of hair in most negroes. More particularly, he asserts, when he comes to discuss the logic of psychological inquiries, that the apparent uniformities discerned in human nature can never be any more than mere 'approximate generalizations,' upon whose lack of certainty he has so strongly insisted: 'as a scientific proposition' we can only assert that 'bodily strength *tends* to make men cour-

[1] See below, pp. 218 f.
[2] *System of Logic*, Bk. III, chap. xxii, § 2, pp. 106 *et seq.*
[3] *Ibid.*, Bk. III, chap. v, § 5, p. 377 (Mill's italics).
[4] *Ibid.*, p. 109.

ageous, not that it always makes them so.'[1] The statistician, of course, would treat the coexistence of such properties as a problem in association or correlation. Since in the case of character-qualities the coexistence is not demonstrably invariable, he would formulate the 'approximate generalization' as an association or correlation with a specific probability; and the group, class, or kind naturally exhibiting these tendencies towards coexisting properties would be defined by a 'factor.'

Most logicians, I take it, would accept the latter part of Mill's statement more or less as I have summarized it. The doctrine of 'Natural Kinds' might not be approved in precisely the same form; but, as I shall point out in a moment, all seem agreed that some very similar postulate is required to validate inductive inference in this field.[2] And, though Mill himself believes that many of these tendencies to coexistence may ultimately turn out to be "an effect depending upon causes," nevertheless even he maintains that "at least some of them" must be "ultimate properties of the 'Kind'."

This, it would seem, must be peculiarly true of 'coexistences' in the psychological field. For why should the human mind be characterized by cognitive, affective, and conative tendencies at all? Why should grass look green rather than (say) feel pink or smell cold? Why should this physiological condition be accompanied by that particular feeling or by that particular desire rather than by some other? In all such cases we can certainly see no necessary connexion, either direct or mediate, between the properties thus correlated; and in many of the cases it is surely out of the question that we can ever demonstrate that the connexion is in any way necessary. It may be a material impossibility, but it is obviously not a logical impossibility, that the stimulation of a particular type of retinal or cerebral cell should be accompanied by a red sensation rather than by a green, just as it was never logically impossible that crows may be white, negroes have straight hair, or oxygen be inflammable.[3]

If we accept this standpoint, it would seem to follow that, by the formal arguments of factor-analysis alone, the psychologist can

[1] *Loc. cit.*, Bk. VI, chap. v, § 4, p. 446. These 'empirical laws of human nature,' he adds, 'must at any rate in part rest on causal laws of the formation of character: but these "cannot be ascertained by observation and experiment, but must be studied deductively."'

[2] See below, p. 224.

[3] Mill's instances, as he himself points out, are not all on the same footing: some unrealized possibilities may be merely due to the absence of an appropriate cause.

never prove or discover causes as such. What is more, we are led to ask whether the validity of factor-analysis need in any way presuppose the principle of causal determination. In themselves our correlations and our saturation coefficients state only ' material implications,' not ' necessary entailments ' ; they can tell us only that x and y go together as a matter of fact, not that y is impossible and unthinkable without x. Such coefficients, therefore, cannot offer explanations ; they can only give descriptions—descriptions in the first instance of the sample observed, and (if the inductive nature of the argument can be justified) descriptions of the population sampled. Certainly, the causal language used by psychological factorists suggests that they believe themselves to be proving the former type of proposition—namely, explanatory or necessary laws ; nevertheless, it seems clear that, on the basis of factor-analysis alone, they are not entitled to go beyond the latter—namely, mere empirical generalizations of fact. Whether they could ever justify some deeper inference by seeking evidence outside the mere factorial results, and, if so, what kind of evidence would be needed for such a purpose, are problems we must take up later, after we have examined the methods of factorization a little more closely.

Our preliminary review, then, of the uses to which factors have been mainly put leads us to the following conclusions. The logical presuppositions of factor-analysis afford no *prima facie* grounds for treating the resulting factors as causal entities. Even as bases for prediction and inductive inference their value is problematic. Their primary use is descriptive merely ; and only after we have determined what precisely it is they describe can we decide whether they have an inferential and possibly a causal significance as well.

Having considered the chief purposes for which factors may be employed, let us now turn to examine in greater detail the nature of the analytic procedure itself.

CHAPTER III

THE GENERAL NATURE OF THE FACTORIAL TECHNIQUE

The Fundamental Factor Equation.—All the different factorial procedures are derived from the same initial postulate; and the tendency to reify 'factors,' natural enough in itself, receives a silent sanction from the very form in which this initial postulate is almost always presented. Practically every factorist starts with an equation which depicts the mark or score obtained by any given individual in any given test as the *sum* of that individual's mental abilities or 'factors,' each factor being weighted according to its influence on the particular process tested. Thus "the tester," to take Thomson's illustration, "hopes to give the composition of his test as

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

where g is Spearman's g (intelligence), v the verbal factor, n a number factor, and s the remaining specific of the test; and the coefficients are the 'saturations,' i.e. the correlations believed to exist between the test and those fictitious tests called factors, the squares of the saturations (factor-variances) adding up to unity."[1] Thurstone,[2]

[1] *Loc. cit.*, p. 18: (I have slightly condensed Thomson's wording). In this paragraph he is showing how Spearman's theorem may be extended to include group-factors. But Thomson's own equation has the same form (cf., for example, *J. Educ. Psychol.*, XXVI, p. 242, eq. [1]); Spearman's two-factor equation would put zero for the second and third coefficients and suitably adjust the first and last.

[2] [84], p. 52, eq. [1]. Cf. [122], pp. 2–3, "The first simplifying assumption of the factorial methods is that the performance of a task . . . can be regarded as a sum of the contribution of two (or more) primary abilities," or, in other words, that it "can be expressed, in a first approximation, as a linear function of these primaries. . . . If we know . . . the weights and . . . the scores in the fundamental abilities then the objective performance can be predicted." May I add that the theoretical chapters of this little monograph give an admirably lucid account of the problem of factor-analysis and the 'centroid

Hotelling,[1] and Kelley[2] all begin with a similar equation. Thurstone writes it in more general terms, $s_{j1} = a_{j1}\,x_{1i} + a_{j2}\,x_{2i} + \ldots + a_{jq}\,x_{qi}$, where s_{ji} is i's score in test j, x is i's measurements in the 'q statistically independent reference abilities,' and a the 'factor loadings' or saturation coefficients. Using matrix notation,[3] we might express it still more succinctly as $M = FP$, where M denotes the empirical measurements, F the factor loadings, and P the hypothetical factor-measurements for the population tested.

Now the form of all these equations suggests that the observable capacity, as empirically tested and measured, is composed of four or more fundamental 'abilities' or 'factors' added together, just as the value of 10 dollars can be obtained by adding 2 pounds to 2 shillings, 7 pence, and one halfpenny. I myself, however, have argued that " to begin with an equation like $M = FP$, when P is not given and M is, seems highly illogical: it is far more natural to start from the equation $P = WM$," i.e. deduce the hypothetical factor-measurements as a weighted average of the observed test-measurements instead of *vice versa* ([101], p. 84). This would emphasize from the very outset the literal truth, namely, that the factors, not the test-performances, are the hypothetical quantities to be obtained by algebraic summation. After all, almost every factorist, who gets so far as to tell us how he would calculate his factor-measurements, writes out a regression equation conforming to this second type,[4] though, instead of making it the starting-point of his exposition, he usually appends it as an afterthought.

Factors as Averages.—This alternative equation, expressing the hypothetical factors in terms of the empirical obser-

method' for those who feel unequal to the fuller and more technical presentation in *Vectors of the Mind*?

[1] [79], p. 418, eq. [3].

[2] [85], pp. 2 and 60.

[3] Those who are not familiar with matrix notation will find illustrative examples of this and the following equations worked out in full in Appendix II.

[4] E.g. Thurstone [84], p. 226; Thomson [132], p. 115; Spearman [56], p. xviii; Kelley [85], pp. 58–61.

vations, brings home to us what is constantly forgotten,[1] namely, that by its very mode of computation *a factor is simply an average or sum total of certain measurements empirically obtained.* To estimate a factor-measurement nothing more mysterious is involved than the process of averaging, with or without appropriate weights.

In the case of the first or general factor, where all the weights are positive, this conclusion is obvious enough. Every examiner who wishes to estimate the general ability of his candidates takes the plain unweighted total or average of the marks they obtain in the papers he has set ; and even the professional factorist will in actual practice seldom stop to calculate the various weighting coefficients that his full regression equation requires, but will similarly compute the unweighted average or the plain arithmetical sum. Factor-analysis merely makes this familiar procedure a little more precise by introducing a differential weighting.

In the case of the more specialized or ' secondary ' factors the factor-measurement is the sum or average of the standardized deviations about the first factor (or the factor last calculated), duly weighted if we desire to be precise. This perhaps is not so obvious: but I shall endeavour to demonstrate it in a later chapter.[2] To put it in its simplest terms, let us suppose that we desire to estimate a child's verbal ability ; we must apply verbal tests, and, since these are inevitably influenced by general intelligence as well, we must also apply tests of intelligence : to get a rough approximation to his verbal ability,all we have to do is first to deduct from his performance in the verbal tests whatever may seem attributable to his general intelligence, and then—since weighting rarely makes much difference to the average—we may take the unweighted average of the residues, provided,

[1] When first setting down this view, I imagined I was merely putting into words what every factorist probably thought too obvious to need explicit statement. Stephenson, however, has strongly criticized the notion, which he considers to be " quite contrary to the tenets of the Spearman school (of factor-analysis)." " A factor," he insists, " should be clearly distinguished from a mere average " ([98], p. 357). And, as we shall find later on, he holds that this erroneous assumption invalidates my conception of the factors to be obtained by correlating persons.

[2] See below, Part III, p. 399.

of course, that the test-results are expressed in comparable units.

But the actual details of the calculations need not trouble us for the moment. My point is that, whether we are concerned with primary or secondary factors, the factor itself always appears at the end in the character of an average or sum, that is (*prima facie*, at any rate), as a synthetic rather than an analytic result: and the purpose of the so-called 'analysis' is simply to find the best possible ways of grouping the available tests so as to represent this general characteristic or that, and the best possible weights to give to each test, when we require something more than a rough approximation. It is during this preliminary sorting that the procedure takes on a more technical and, as we shall see, more controversial appearance.

Factors as Patterns.—The simplest weights available are + 1, 0, and — 1. If we introduce no further differentiation, we can still treat our factors as averages to be obtained by simple addition.

Consider, for example, the following set of traits extracted from a rating-scheme, where pupils of school-leaving age were marked on a standardized bipolar scale for a list of characteristics, including the chief school subjects and McDougall's 'primary emotions.'

Traits.	Factors.			
	General Intelligence.	General Emotionality.	Introversion.	Cheerfulness.
English Composition	+ 1	0	0	0
Problem Arithmetic .	+ 1	0	0	0
Sociability . .	0	+ 1	— 1	+ 1
Anger . . .	0	+ 1	— 1	— 1
Tenderness . .	0	+ 1	+ 1	+ 1
Fear . . .	0	+ 1	+ 1	— 1

In a previous investigation on vocational guidance it had been found that the four 'key qualities' which (*a*) were of greatest practical importance, and (*b*) accounted for the greatest amount of variance in a set of character-studies sent in by teachers, social workers, and works' managers, were the 'general' factors of intelligence and

emotionality and the ' bipolar ' factors of introversion and cheerfulness. Now for these four key qualities very good estimates can be obtained by simply weighting the foregoing six traits as indicated in the body of the table : for the general factors, the marks for the intellectual and the emotional traits are to be added just as they stand ; for the bipolar factors, the marks for some of the traits have first to be reversed or ' reflected ' before they are added, as indicated by the negative weights.

Since, however, the correlations between such factors and such traits—the ' saturations,' as they are commonly called—are never absolutely perfect or absolutely zero, a somewhat better assessment could evidently be obtained by substituting fractional weights for the simple plus or minus signs. Even performances in tests of Composition and Arithmetic are not wholly unrelated to the temperamental factors, the introverts being slightly better, and the extraverts and the emotional pupils being slightly worse, at Arithmetic than at Composition. In practice fractional weightings would be almost essential, if we were aiming at a more precise and detailed specification of the pupils' cognitive qualities—e.g. if our estimates of intelligence were to include proficiency in less academic subjects (such as handwork and drawing, for example) or if we desired to distinguish between pupils fitted for technical and clerical vocations respectively. But, whether fractional or not, such weightings obviously enable us to effect an economy, not so much in the number of hypothetical factors, as in the number of actual tests.

Here the attitude of the practical psychologist shows a curious contrast to that of the theoretical investigator. The theoretical investigator wants to describe a maximum number of tests in terms of a minimum number of factors. The practical psychologist would rather aim at *deducing a maximum number of factors from a minimum number of tests*. Thus, in a vocational inquiry it was found possible to deduce from six tests only—Composition, Spelling, Problem Arithmetic, Mechanical Arithmetic, Drawing, and Manual Dexterity—ten factors of vocational importance, all more or less independent of one another, and so to classify the examinees (no doubt, very roughly) under ten different heads, namely, Intelligent, Verbal, Mathematical, Technical, Artistic, Logical, Accurate, Quick, Emotional, Introverted.[1]

[1] The object of the inquiry was, quite tentatively, to see how far it might be possible to reduce the time required at vocational guidance examinations by devising tests that should facilitate estimates for as large a number of ' factors ' or ' key qualities ' as possible. The classification

In all such cases, as may be seen from the table above, each factor is characterized by what may be called a vectorial or *columnar pattern*. When the weightings are fractional, each is distinguished by a unique series of fractional figures for a certain relevant set of traits, that is to say, by the saturation coefficients for those traits. And each factor differs from every other, not merely by the set of traits to which it is positively related, but also by the relations between the trait-weightings themselves.

This is no new conception. Nearly all biological classification rests on the same principle. We put all birds in a class together, separate from mammals, fishes, and amphibians, not in virtue of some single and simple avian 'essence'—their capacity for flight, for example—but because their limbs and organs are differently related as regards shape, size, and position from those of other vertebrates—in a word, in virtue of their special mode of *organization*—their characteristic structure, viewed as a whole but with special emphasis on crucial points. The relations beween their limbs and organs thus constitute a general diagnostic pattern that defines the type.

The notion of a complex and distinctive attribute that is expressed by a number of separate characteristics, and yet is to be thought of as a unitary whole, is one that should present no difficulty to the modern psychologist. The word *Gestalt* expresses just this conception. When the attribute is a mental factor, specified in quantitative terms, the nature of the synthesis can be exhibited in graphic form. If each contributory trait is represented by a vertical line,

deduced was checked by direct assessments made by schoolmasters. The master's assessments for the different factors were by no means wholly uncorrelated with each other. Moreover, in the case of the temperamental factors the agreement between the master's assessments and the classifications deduced from the tests was low. The results, however, were sufficient to show that, from a practical standpoint, the construction of tests, which, like the Binet scale, enable rough judgments to be made for a number of different factors at a single examination, would lead to much saving of time and labour. Such a conclusion seems fully in keeping with Fisher's view that, in experimental work, 'excessive stress is often laid on the supposed importance of varying only one factor at a time' ([109], pp. 100 *et seq.*, § 37, 'The Single Factor').

all on the same horizontal base, and if the length of each line is proportional to the trait-measurement or weight, then the special features of the factor will be outlined by the pattern or contour which the tops of the verticals present. Diagrams of this kind are familiar enough to the practical psychologist. Long before the advent of formal factorization, he was accustomed to charting the mental constitution of individuals in the form of a 'psychogram' or a 'profile,' and to labelling each person as belonging to this type or that according to the shape of his curve. The additional contribution that the statistical psychologist can make is to increase the accuracy of the diagnosis by formulating standard 'curves' to describe the pure or ideal type.[1] If our assignments are to be exact, it is not enough to compare contours by eye; the specifications both of the type and of the individual must be based on figures rather than on diagrams, and each person's approximation to this type or to that can then be stated as a correlation. Such a summarized statement is precisely what factor-analysis, with its saturations and regression coefficients, sets out to give.

Much the same principle underlies the mathematician's suggestion that the factor can be represented by a vector; for a vector is defined by a set of scalar numbers, stated in a definite order, and called its co-ordinates. In using co-ordinates instead of ordinates, he implies that the lines whose lengths represent the numbers or weights are to be drawn in different directions and to start from a common origin instead of standing parallel on a common base line. It is as though we saw three pins of different length, lying side by side on the pin-paper, and then took them out and stuck them round a pin-cushion instead; the two little patterns formed by the pin-heads—the old 'curve' on the flat paper and the new three-dimensional pattern round the cushion—would be two alternative ways of representing the same complex fact, namely, the differences in length of the

[1] Thus, illustrative diagrams for typical children high and low in the general factor of intelligence (or rather general educational ability) and in the special verbal and arithmetic factors respectively were given in my 1917 L.C.C. *Report*, pp. 64–5. Cf. also Figure 1 below, Part III, p. 427.

pins. With n traits instead of 3, the vectorial pattern will be specified by co-ordinates in n dimensions: but, if the student has trouble over picturing patterns in n-dimensional space, I suggest he translates the n-dimensional configurations into flat zigzag 'profiles' by treating the co-ordinates as ordinates.

Factors as Axes of Co-ordinates.—When, however, we try to enlarge not only the list of traits but also the sample of persons, when, that is to say, we have to deal simultaneously, not with one or two individuals only but with a very large number, and possibly with a number of different types to suit the different groups, then some form of n-dimensional representation, or at least the language of n-dimensional representation, becomes inevitable. We start with N persons tested and measured for n correlated traits; and we desire to convert this empirical mark-sheet into a more convenient form by describing the same N persons in terms of r uncorrelated factors. How is the pattern of weights to be deduced?

When the problem is put in this generalized fashion, the mathematical arguments take on a somewhat formidable aspect. For the elementary student the simplest mode of exposition is to outline the proof for two or three variables only; and then show that the number of variables is really irrelevant to the argument so that the proof can be generalized for any number. But if matrix algebra is used, the proofs for variable matrices are almost as simple as the proofs for variable scalars.[1] This was the procedure adopted in my early *Notes* [93]; and algebraic proofs need not be repeated here. In general the transformations employed follow the methods regularly used in elementary geometry for translating measurements obtained in terms of one set of co-ordinates, presumably a provisional or casual set, into terms of a second set, chosen so as to be

[1] Nearly all Spearman's proofs relating to a single factor, for example, can be generalized in this way. It is curious that Thurstone, after his lucid introductory exposition of matrix algebra, uses it so little in his subsequent proofs, relying instead almost entirely on the old summation notation: (cf., for example, the simplicity of the matrix proof of the formulæ for 'appraisal of abilities' with that given by him in [84], chap. x).

better fitted for permanent reference, which usually means reference to a standard system of orthogonal axes, at once independent of each other and of the accidental circumstances of the observations.

Although the introduction of the idea of resolving observed tendencies into independent components seems to have struck the psychological worker as something wholly novel—so novel in fact that he often supposes the whole procedure to be an invention peculiar to psychology, and to a special school of psychology at that—nevertheless there is scarcely a branch of science, pure or applied, into which such transformations do not constantly enter. I have already cited one instance from geography. Let us glance at another, almost equally familiar and far more instructive. The theoretical psychologist who dismisses the introduction of mathematics as far-fetched, and the applied psychologist who regards arithmetical computations as an unpractical hobby of the faddist, will do well to consider how similar mathematical and arithmetical devices have become part of the everyday tasks, not only of scientific investigators, but of practical workers in numerous other fields.

A navigating officer on the high seas will measure the ' altitude ' and the ' bearing ' of a particular star, i.e. its apparent height above his horizon and its apparent distance east or west of his meridian. If another observer measured the apparent position of the same star simultaneously from another spot, he would obtain a different set of measurements ; yet, if compared, the two sets of measurements would evidently be related to each other in a way which would not have arisen had each observer measured a different star. The navigating officer therefore performs a routine calculation, and converts his measurements into standard terms, namely, the distance of the star above or below the celestial equator (or equinoctial) and the distance east or west of the vernal equinox (first point of Aries), thus finding its celestial *latitude* and *longitude*, or, as he would call it, its ' declination ' and ' right ascension ' : these will be the same at all times and places, and will have the further advantage that, for different stars, the specifications will now be quite independent of each other.

A navigator of the Spearman school, more interested in theory than in practice, might perhaps prefer to express his results in terms

of three axes at right angles to one another drawn through the centre of the earth—a central axis through the poles, which would represent the appropriate *g*, and two supplementary or 'specific' axes orthogonal to this, through Greenwich and through Galapagos. An astronomer of the Thurstone persuasion would reply that it was uneconomical to use three sets of co-ordinates where two would suffice, and further that the last two axes had 'no astronomical significance' and so must be rotated. He would doubtless elect to substitute the plane of the earth's orbit (the ecliptic) and the plane of its equator, thus rotating the results of the initial analysis and referring the entire data to a couple of oblique planes or axes only.

But throughout all the arguments and calculations, the general formulæ which the navigator employs are almost identical with those which the psychologist would use, if the heavenly bodies were a set of persons to be tested, and if their measurements by different observers at arbitrary spots represented the performances of the persons with so many sets of arbitrary tests. Thus, every day of his voyage, the navigator is carrying out a 'factor-analysis.'[1]

We may push the analogy further. So far we have spoken of our observers as interested primarily in the positions of the stars; but they will adopt exactly the same methods to determine the positions of their ships. Similarly the psychologist can use the same observations and the same mode of analysis to compare either the characteristics of the individuals he is testing or the characteristics of the tests he employs. With either mode of approach the navigator's task is doubtless far simpler: for he is concerned with two dimensions only—latitude and longitude. But evidently the

[1] The analogy leads me to point out in passing the occasional advantages of using spherical co-ordinates instead of the more usual Cartesian co-ordinates. As I have elsewhere shown ([93], pp. 247, 300), the formulæ used in multiple factor-analysis are merely special applications of the ordinary formulæ for multiple and partial correlation; and for three variables the latter were given a suggestive interpretation in terms of spherical trigonometry by Karl Pearson [10]: the procedure thus indicated can be generalized to any number of dimensions. For teaching purposes, three factors and their rotations can be represented very vividly to the eye by plotting coefficients on a black globe, such as is used in classes on geography and navigation. Even non-mathematical students have usually picked up some elementary notion of spherical geometry from their early lessons on geography; and for them the rudiments of factor-analysis can be expounded quite simply in this concrete form. The approach by spherical trigonometry was used in one or two early factorial papers, and will be found to lend itself to interesting theoretical investigations, suggestive alike of new formulæ and new working methods.

principle may be generalized. If, for example, we were piloting, not a surface vessel but a submarine or an aeroplane, he would require to work with three dimensions. And Jules Verne, who imagined a vessel steering through a universe of n dimensions, had merely to set his navigator the same problem in n-dimensional form. As a matter of fact, numerous questions in aeronautics are nowadays solved by expressing the spatial data as matrices, and then reducing the matrices to terms of simpler factors by precisely the same iterative devices as are used for finding factor saturations in psychology.[1]

These analogies suggest a further conclusion. They create a strong presumption that the factors in terms of which the psychologist ultimately expresses his results can at most claim only the same kind of existence as the lines or points to which the navigator refers his measurements. The ecliptic and the equator, the poles and the first point of Aries, are not concrete objects like the stars themselves: they are simply items in an *abstract frame of reference*. As such they are naturally presumed to be constant; but they are wanted merely for descriptive purposes; and no one would be tempted to assign them an actual physical existence. No doubt, where selection is possible and knowledge permits, it will be convenient to choose such points and lines as possess some simple relation to extraneous bodies.[2] When, however, such extraneous existents, and relations to such existents, are to be established, something more than a mere mathematical analysis is employed.

[1] The interested reader will find it instructive to compare some of the methods employed by the Aerodynamics Department of the National Physical Laboratory (see, for example, the recent volume by Frazer, Duncan, and Collar on *Elementary Matrices and Some Applications to Dynamics*, Cambridge University Press, 1938, especially the worked examples of 'iterative numerical solutions of linear dynamical problems,' pp. 133–154 and 308 *et seq.*).

[2] Thus, as I have argued elsewhere, it might be of advantage to define the leading factors in psychology by some provisional convention that is more or less arbitrary, much as we define the zero meridian or the celestial north pole. Until we know more of the functional relationships that obtain between different mental characteristics, this would probably be the most convenient procedure in actual practice. Yet it is not strictly necessary in theory. As more advanced students will be aware, a calculus has been devised in pure mathematics which enables the physicist to perform his analysis of space or of space-time in a way which leaves the axes of co-ordinates entirely undetermined: he is thus able to describe a complex pattern of relations without having to specify what are the relata. *Prima facie*, therefore, there is no real need for the psychologist to seek stability for his factors by forthwith endeavouring to identify each one with some concrete psychological reality, as is so commonly supposed.

Those who regard factor-analysis as the hobby of a special school, which the ordinary student, with no taste for numbers, can safely ignore, may be reminded that almost exactly the same procedure has found a reference in nearly every psychological textbook ever since textbooks began to incorporate experimental work. For the ordinary student of psychology, perhaps the simplest, earliest, and most familiar examples of this kind of analysis are (i) Wundt's attempt to determine 'the number of elementary feelings' and (ii) the attempts of Helmholtz and other experimentalists to determine the number of elementary or 'primary' colours. Though the phrase was not explicitly used, the object of such efforts was to reduce the phenomena of feeling and of colour vision to terms of 'orthogonal factors.' Wundt, it will be remembered, "in developing a comprehensive theory of feeling, postulated three dimensions: (1) excitement-quiet ('depression'), (2) tension-relief, (3) pleasantness-unpleasantness. The total feeling at any moment can be located by reference to each of these dimensions, just as a point on the earth can be identified by latitude, longitude, and altitude." [1] It is interesting to note that the nature and independence of these three 'dimensions' are in some measure confirmed by recent endeavours to analyse the complex temperamental characteristics of individuals, by means of a modernized factorial technique,[2] though Wundt himself based his analysis rather on introspection and the study of emotional expression than upon a formal mathematical analysis.

The determination of the 'laws' of colour mixture has a

[1] Wundt, *Grundriss der Psychologie* (1896), pp. 98 f. Woodworth, *Experimental Psychology* (1938), pp. 235 f.

[2] In particular by the statistical evidence for excitable, repressed, and cheerful 'types' or tendencies and the reverse: cf. [114], [129]. Wundt's own analysis of temperamental types was two-dimensional rather than tri-dimensional; but his tri-dimensional theory of the feelings seems to have had considerable influence on later German attempts to classify temperaments (cf. [21], pp. 188, 483 f.). Wundt's pupil, Titchener, not to mention many other critics, has adduced strong reasons for doubting whether the "dimensions are really 'independent' (as naïvely understood)" (*Phil. Stud.*, XX, pp. 382 f.); yet until recently no attempt seems to have been made to test the 'independence' by the obvious method of correlation.

more rigid mathematical basis. By examining the relations and resemblances of visual sensations, obtained not from mere introspective comparisons but from quantitative experiments on the effects of mixing lights of varying hue, it is possible to secure a series of colour-equations ; and it is then not difficult to show that these equations are reducible to terms of three independent variables only—the so-called primary colours, three colour-factors that can be represented by the three dimensions of the familiar double colour pyramid. We may even determine what three primary colours will best fit the actual data.[1] But whether there are really three separate retinal substances or processes corresponding precisely to these theoretical primaries is a question that cannot be decided from the colour equations alone.

The mental ' factors ' with which recent factor-analysis is more commonly concerned—the ' primary abilities ' deduced from tests applied to school children or adults—have precisely the same abstract nature. Like the primary colours deduced from the colour equations, they are postulated to provide a standard frame of reference. They are, in short, as the schoolboy is taught to say, ' component vectors ' into which ' resultants ' may be ideally ' resolved.'

Analogy between Factor-analysis and the Resolution of Forces.—The foregoing illustrations, however, bring to the fore an important point of difference between factor-analysis in psychology and analogous methods of solving geometric problems in the simpler physical sciences. The dimensions with which the navigator or the aeronaut is primarily concerned are directions in actual space : the dimensions with which the psychologist deals are directions in a diagram only, and the lines of which he speaks represent changes in non-spatial variables, not differences in actual length or actual

[1] In our own laboratory, Mr. P. H. Chatterji has recently been working over the problem of colour-vision afresh from this point of view, applying a more up-to-date factorial technique both to his own data and to that recorded in the literature. I may add that many of the time-honoured problems in sensation, hitherto attacked by so-called psycho-physical methods only, could be fruitfully taken up anew from the standpoint of the analysis of variance or covariance.

direction. They are like the lines on the temperature chart that hangs above a hospital bed, where the base line indicates the ' length ' of the patient's illness, and the vertical lines, measured upwards from the base, indicate the ' height ' of his temperature, or like the rectangular lattice of lines on the record of a self-registering barometer, where a horizontal movement marks the passage of time and a vertical movement marks the rise and fall of atmospheric pressure. Moreover, except in the simplest cases, the factorist cannot actually plot his graphs : for, as we have just seen, they would run into far more than two or three dimensions. What excuse, then, can he claim for using spatial terms and introducing geometrical or trigonometrical concepts, and that not merely to describe or represent, but apparently to analyse and simplify, measurements of such insubstantial attributes as mental abilities or temperamental traits ?

The mathematician doubtless would be content to reply that the two alternative methods of analysis—algebraic and geometrical—are ' abstractly identical ' : [1] we have only (he would say) to *define* a ' point ' as a set of n real numbers and then consider the class of all such points, and geometry is turned into algebra (cf. [81], p. 1) ; conversely, we may agree to call the n real numbers a ' point,' and algebraic arguments can be expressed in geometrical language. But to the student unfamiliar with the logic, and (I think we should add) the history, of mathematics, this short and sweeping answer is scarcely convincing : such an abrupt identification bewilders rather than helps him. He looks upon his geometrical diagram as an aid to visualization, not as an analytic device : for him, the numbers are real, and the lines are pictorial symbols ; whereas, as I shall try to show in a moment, when we are dealing with mental qualities, the numbers themselves are also symbols to assist our argument. To the theoretical investigator, on the other

[1] " The whole content of metric euclidean geometry of any number of dimensions is contained in the implications of the assumptions defining the ordinary algebra of real numbers. The two procedures are therefore abstractly identical : their difference consists merely in the arrangement of the logical sequence." (J. W. Young, *Fundamental Concepts of Algebra and Geometry*, pp. 182–3.)

hand, the avowed identification of the abstract methods may lead to an unconscious and unwarranted identification of the concrete subjects: and thus, though the processes designated by the numbers or the points are no longer material but mental, the spatial terminology is still apt to suggest a spatial and physical interpretation. Mental factors are described as 'mental forces' or as 'mental energy'; and these phrases in their turn are subsequently treated, not as bold metaphors, but as implying something fundamentally kinetic.

It would appear, therefore, urgently desirable to trace out the correspondence in clearer and more explicit terms. Although in themselves non-spatial, mental conceptions, I suggest, may nevertheless be legitimately given a spatial representation if we accept the following principles. The initial steps, it will be observed, are analogous to those adopted in other sciences—in thermo-dynamics, for example —where the language of analytic geometry is used in formulating somewhat similar problems.

(i) The first step involves no difficulty. A point at the end of a line, having a fixed origin and a specified direction, may be used to represent, not only an actual line having the same relative direction and the same proportionate length, but also motion along that line, and (if the masses moved and the time of the movements are assumed to be the same) the *forces* implied by[1] such movements. Thus a line on a map might be used to indicate, not only the distance from Dover to Calais, but also the force required to move a vessel

[1] I do not say the 'forces *causing* such movements.' By a force I understand, not a perceptible push or pull (such as we ourselves might feel in a tug of war), but a certain abstract algebraic function of mass and motion, lending itself to convenient statements of equality (as in formulating the conditions for two or more poised weights in equilibrium). Thus, instead of talking of such and such a movement as produced by such and such a 'cause,' it would be better to say that the movement appears as the necessary logical consequence of some conceivable change that is logically prior. The point is important in psychology, because of the popular associations still clinging to the terms force and energy: these lead the student to think such phrases as 'mental force' or 'energy' refer to a perceptible effort or exertion; whereas, at any rate in the present context (cognitive testing), they refer merely to the relative efficiency of the mental process.

from Dover to Calais. Now, we have seen that it is desirable to refer such movements and such forces to a standard frame of reference; and, as every schoolboy learns in the 'parallelogram of forces,' mechanical force may always in theory be resolved, in accordance with the cosine law, into independent components represented by lines at right angles: e.g. if the vessel is moving south-east at x miles per hour, this may be due to a wind blowing east at ($x \cos 45°$) miles per hour and a current running south at ($x \cos 45°$) miles per hour, or perhaps to what we call (summing up an implied resolution in a phrase) a south-east wind.

(ii) The next step is to use similar points to mark differences in physical state *other* than those of mere difference of position—e.g. differences in heat, pressure, volume, chemical composition, electric potential, and the like. The lengths of the lines will then be proportional to the amount of change, and the differences of direction will mark differences in the *quality* of the change, not differences of direction in actual space. Thus, one line may represent increase in pressure and another increase in volume, and we may seek to explain these as the inevitable accompaniments of a concomitant rise in temperature. With this further extension in the significance of the symbolic lines, there is no longer any need to restrict ourselves to three independent dimensions: we can have as many as we like, though, of course, it will then be impossible to represent all of them at once on flat paper or by a model. At the same time, however, the whole system of quantitative changes may be regarded as the result of a transfer, or rather of a transformation, of one underlying capacity, constant in amount, but capable of being applied in many different directions—namely, energy.

(iii) Having generalized so far, the transition from physical processes to mental is easy. We may think of the application of a test to a testee as the disturbance in the equilibrium of a mental system, leading to a progressive change of state in that mental system—such and such an extent of change being visibly registered during a unit of time, provided the conditions are kept constant. In every case the essential nature of the process can be described by saying (with a natural expansion of the strict meaning

of the terms) that work is done against resistance—i.e. by overcoming the difficulty of the task.

The differences between the various tests, however, are qualitative differences. Consequently, we must devise some conventional rule for expressing qualitative differences quantitatively. Let us assume that a similarity in the direction of two lines denotes a corresponding similarity in the nature of two tests. We can then use the calculated correlation between those tests to measure the agreement in direction: coincidence of direction will represent perfect correlation; orthogonal directions will represent zero correlations.[1] With these assumptions we may now legitimately enlarge the notion of directed forces, and, if we like, speak of '*mental* forces' as responsible for these mental changes of state. On this basis we can regard any given test-performance as the resultant of hypothetical component forces or 'factors,' so chosen as to be mutually independent; and the correlation between any two tests will then be deducible from the correlations between those two tests and the components or factors in accordance with the familiar cosine law.[2]

[1] In the theory of factor-analysis, the idea of expressing correlations as angular functions was, I think, first mooted in connexion with emotional tendencies [30]. The notion was based on Pearson's interpretation [10] of partial correlation in terms of spherical trigonometry, referred to above: taking the coincidence of directions ($\theta = 0°$) to represent perfect agreement, reversal of direction ($\theta = 180°$) to represent perfect disagreement, and therefore $\theta = 90°$ to represent zero agreement, the natural functions are $r = \frac{1}{90}(90 - \theta)$ or $sin\ (90 - \theta)$. On this basis it was suggested that correlations could be "represented inversely by distances of arc" ([30], p. 696 and diagram). In re-examining some of Webb's data and my own, Maxwell Garnett gave a better and more formal expression to these vague suggestions: extending the familiar correlation diagram to n dimensions, he formally deduced what he called the applicability of the cosine law: putting $\theta = cos^{-1} r$, we may write $cos\ \theta = cos\ a_1\ cos\ b_1 + \ldots + cos\ n_1\ cos\ n_2$, where $cos\ n_1$ denotes the correlation of the test 1 with the nth factor [37].

An equally legitimate, but entirely different procedure (or so it might seem at first sight) consists in expressing the correlations, not as cosines of angles, but as ratios of the axes of the frequency-ellipse or ellipsoids: the relation between the two alternative modes of representation is obvious enough on considering the ordinary correlation diagram for two variables only, and was given a general formulation in [93], pp. 253 *et seq.*

[2] See below, p. 91.

(iv) We can, if we think it helpful, take yet another step, and suppose that both the test performances, and the forces in terms of which they are to be described, are manifestations of one and the same capacity for work, which can expend itself in different directions. This single underlying capacity it will be natural to call ' energy.' We can then say that different tests, according to their differing difficulty, will require different amounts of the same ' *mental energy*,' irrespective of their specific quality or kind ; and we can go on to declare that different individuals must possess a different amount of ' mental power,' defining power in the usual way as the rate at which energy is expended, and measuring it in terms of the amount of work accomplished per unit of time.

But it now becomes evident, I think, that the energy which we thus postulate is a purely logical construction, and has no more than a conventional significance : we can hardly treat it as " in the last resort presumably identifiable with the neural, neuro-muscular, or neuro-endocrino-muscular energy of the organism," or with the " energy that serves the whole cortex or possibly the whole nervous system" (cf. Spearman [56], pp. 121 f.). Certainly, its amount cannot be identified with, or even directly related to, the amount of physico-chemical energy required for the propagation of nerve-currents. No doubt, during mental work the heat production of the central nervous system (exceedingly minute in any case) is somewhat increased ; but, so far as we can discover, the increase is at least as great during unintelligent physical activity, and probably far greater during emotional excitement. If we include other sources of physiological energy—the muscles and the endocrine glands—the lack of correspondence between physical work and mental work is still more flagrant.[1] All these physiological forms of energy have their mechanical equivalent :

[1] In such forms of work as sawing wood, the extra calories per hour may rise to 200 or 300. But with " strong mental effort expended in solving mathematical problems " Benedict found an increase of only 3 or 4 per cent. He adds : " the cloistered scholar at his books may be surprised to learn that the extra calories needed for one hour of intense intellectual effort would be completely met by eating one half of a salted pea-nut."

their amounts can in theory be translated into terms of mass, length, and time. Intellectual output we can express as the production of such and such a change against such and such a resistance within such and such a time : but the changes and the resistances can no longer be literally *described* in terms of length, mass, and time ; they can only be figuratively *represented* in terms of such units.

This will become clear if we ask what further assumptions would have to be made in order that the amounts of mental work accomplished in performing a series of tests could be equated with energy in the literal sense. Can we, for example, treat the marks as generalized co-ordinates representing the results of the generalized forces acting on the system in question, namely, the testee ?[1] Let us extend the principle of the self-registering barometer, and imagine a mechanism working somewhat on the following plan. To represent the n tests let there be n sliding pieces, moving independently along fixed bars that carry a scale. The mass of each piece (m) must be made proportional to the difficulty of the test it represents. Each piece as it moves will register the state of the changing mental system in such a way that the distance (d) of its outer end from zero on the scale will be proportional to the mark obtained in the corresponding tests. The mental work done in all the tests may then be described in terms of the forces required to move the sliding pieces.

[1] It will be seen that my development of the underlying analogy is based on the suggestion that each test performance can, as a first step, be given the dimensions of a velocity or rate—i.e. be measured by extent of change or amount of output per unit of time—and that each test is then to be weighted according to its difficulty. We might also attempt to give it the dimensions of work. But we should then be confronted with the fact that, in mental processes, the velocities apparently vanish with the driving forces, i.e. when equilibrium is restored. Anything comparable to the inertia of masses (which brings acceleration into such prominence in dynamics) apparently plays but a minor part in prolonged mental processes : each isolated problem no doubt requires its own little spurt, or rather starts its own oscillating waves of attentive effort ; but, as a first approximation, we have to assume that the smoothed 'curve of work' is flat. Thus the cyclic modifications which constitute a mental process are perhaps analogous, not so much to the changes considered in dynamics as to those considered in thermo-dynamics.

Let me add, to avoid misunderstanding, that I am here speaking of cognitive processes only. The factor-analysis of conative processes, particularly those that arise in response to emotional or moral situations, leads also to descriptions in terms of independent 'forces.' But once again these forces have the characteristics of relations, rather than of entities or properties

The resolution of these forces could itself be accomplished by the mechanical model, so long as we were concerned with no more than two or three uncorrelated components. Thus, if d_a is a measurement obtained with test a, and if d_g is its ' resolved part ' in terms of the factor g, then $d_g = d_a \cos g_a$, where $\cos g_a = r_{ag}$. In imagination (though, of course, not in any actual model) this can be generalized for more than three components, so that we can think of n correlated test performances as expressible in terms of n uncorrelated forces. And these forces will still obey the cosine law, namely,

$$r_{12} = r_{1a}\, r_{2a} + r_{1b}\, r_{2b} + \ldots + r_{1n}\, r_{2n},$$

where r_{12} denotes the correlation between the two tests, $r_{1a}, \ldots r_{in}$ denote the correlations between the ith test and the n factors, and $\cos^{-1} r_{12}$ indicates the angle between the two lines representing tests 1 and 2; or, in matrix notation, $R = FF'$. This vector equation, as we shall see, lies at the basis of all factor-analysis: its form evidently implies that the effects of the component forces are to be combined according to the principle of weighted summation.

It is clear, however, that the mechanism I have described is a working model only, i.e. the movements, and the energy involved, are merely symbols like the spatial lengths. In fact, the use of the word ' energy ' merely signifies a particular mode of analysing the problem, not the actual mode in which the nervous system operates. It expresses the fact that, to describe any test performance, we must state three things:

(i) the number of problems solved in a given time $\left(\frac{d}{t} = v\right)$;

resident in individual substances. They are not forces in the mind, but forces in a mental field—a field which consists of the relations between objects, on the one hand, and the subject on the other: thus they might be pictured as ' lines of force ' and have many of the formal characteristics attributed to ' lines of force ' in a magnetic or electrical field. In the last resort conative processes could, I think, be successfully described, at any rate to a first approximation, in terms of the redistribution of a system of physico-chemical energies. But this does not mean that the conative ' forces ' are to be directly identified with physico-chemical ' forces.' The two paths of investigation must go a long way before they meet; and to emphasize their relative independence I have ventured to suggest that the two branches should be called by different names: that dealing with gross concepts—' abilities,' ' instincts,' ' drives,' ' wishes ' conscious and unconscious, and disturbance and recovery of equilibrium in the field of overt behaviour—I term ' psycho-dynamics '; that working with more precise neurological concepts—nerve-impulses, neurone-arcs and ' bonds,' and the disturbance and recovery of equilibrium in the cerebral field—I call ' neuro-dynamics.'

(ii) the difficulty of the problems when solved at a given average speed $\left(m \cdot \frac{v}{2}\right)$

(iii) the degree to which the problems solved resemble those of one (or more) ideal or standard types of tests ($r_{ag} = cos\ g_a$).

On this basis, the amount of energy expended could be equated with the work done, and measured [1] by the product (number of tests) $\times$ (difficulty) $= \frac{d}{t} \times m \cdot \frac{v}{2} = \frac{1}{2} mv^2$. But obviously solving problems of a given difficulty is not *literally* equivalent to moving a quantity of matter m at an average velocity $\frac{v}{2}$.

Analogies between the Mathematical Methods of Modern Psychology and of Modern Physics.—So far I have gone for my analogies to classical or traditional dynamics. To the advanced student, however, the general nature of factor-analysis, and of factors as descriptive constants, will become clearer still if he turns to the more recent developments of physical science, where he will discover an astonishingly close parallel. Indeed, one of the most striking features of factor-analysis is this : not only in its general nature, but also in many minor details the peculiar type of mathematical argument which the psychological factorist has developed is almost exactly the same as that which is employed by the quantum physicist in analysing the fundamental constitution of the material world. In both cases, the argument proceeds in terms, not of single variables, but of twofold patterns of variables, expressed numerically as tables of double entry or ' matrices ' ; and the central problem is to reduce such matrices to a standard or ' canonical ' form by calculating their ' latent roots ' and ' latent vectors.' In both cases, too, the characteristic operation

[1] Alternatively, if we assume the mental work is due to the propagation of nerve-impulses, we could imagine a model in which the results could be represented by the flow of currents. Much the same constants would then reappear : we should have to imagine that the successful performance of the test was equivalent to moving a quantity of current $Ct \propto m$ through an average potential difference of $E \propto \frac{1}{2} \frac{d^2}{t^2}$. But no one could claim, even as a first approximation, that this depicted the true relations.

is what I have called weighted summation, that is, the computation of product-sums.

These analogous techniques have been taken over from mathematicians, and developed by psychologists and by physicists in almost complete independence; indeed, during the earlier stages of their work, each was entirely ignorant of the technical methods which the other was adopting. The reader, therefore, may feel tempted to ask whether they may not have been unconsciously driven to apply very much the same devices because the material world and the mental world are, as we know them, very much akin in their ultimate nature, and so yield to the same mode of analysis: both being essentially describable in terms of patterns of relations between unknown relata. That, however, is a question which we must postpone until we take up the metaphysical issues.

Mathematical Arguments in Psychology.—The technical methods that have been thus worked out may be regarded as a special application of a branch of higher mathematics which has received the name of the 'theory of groups' [71]; and, as I shall argue later,[1] it will probably be to the theory of groups itself that the psychologist of the future will turn directly for his analytic technique. Here it will be sufficient to note that, in the last resort, the apparent reason why we can deal with intellectual processes by methods analogous to those used in dealing with kinetic processes is, not that the former, like the latter, are manifestations of one and the same physical energy, but that in both cases the resulting changes can be expressed as the effect of group-operations. Now, the philosopher would consider the theory of groups, not as a branch of mathematics, but as a branch of formal logic; and, indeed, most mathematicians would nowadays agree that their science at its widest is to be regarded, not in the old-fashioned sense as the science of quantity, but as the science of logical relations: 'logic and mathematics differ as boy and man: logic was the youth of mathematics; mathematics is the manhood of logic.'[2] This suggests a conclusion of the utmost

[1] See below, p. 242.

[2] B. Russell, *Introduction to Mathematical Philosophy*, p. 194.

importance : in my view, we should think of *factor-analysis as a logical method rather than as a mathematical method.*

Once this standpoint is adopted, many of the objections that philosophers have so often urged against attempts to apply arithmetical procedures to the mind fall immediately to the ground ; for it now appears that the numbers and other quantitative devices which the psychometrist has introduced are in the first instance merely symbolic expedients, employed to help him to state his arguments in a more precise and rigorous form.

CHAPTER IV

THE LOGICAL STATUS OF MENTAL FACTORS

Logically, Factors are Principles of Classification.—To the ordinary student who is new to psychological research, an attempt to explain factors in terms of analytic geometry is in most cases an attempt to explain *ignotum per ignotius.* In this country, such a student is more likely to have approached psychology from the philosophical than from the mathematical side. Hence he will be able to appreciate the nature of factor-analysis far more easily if he is shown its logical origin rather than its mathematical origin. And, as I have just insisted, its logical nature is, after all, the more fundamental, although too often that nature is obscured instead of elucidated by the mathematical technicalities with which it is commonly expounded.

Let us, then, forget for a moment that the psychologist, in his effort after precision, puts his factorial specifications into numerical terms, and let us imagine that a verbal statement will be sufficient. The so-called factors, as we have seen, are used because we continually need a few, permanent, and pregnant concepts by means of which we can describe *both* persons *and* traits. This double mode of description will, I think, become intelligible enough if we consider once again a concrete example such as those I have cited above.

An unknown child is examined at the clinic. The case-report sums up his physical and his mental characteristics under separate heads. On the mental side, it describes first his intellectual (or cognitive) and then his emotional (or orectic) behaviour. Intellectually he is, let us say, mentally defective as regards general intelligence, verbal as regards his special type; and his educational age in each scholastic test is so and so. Temperamentally, he has high general emotionality, is of repressed or introverted type,

depressed rather than cheerful, displays oversexed and timid propensities, but is not ill-tempered or self-assertive; and so on. In all such descriptive summaries we rely primarily though not exclusively on a few key qualities to which factor-analysis (or an intelligent anticipation of factor-analysis) has led us—general intelligence, general emotionality, specialized 'abilities,' specialized 'temperaments,' and perhaps the primary emotions or 'instincts.' These are all attributes *from which numerous secondary characteristics can as a rule be safely deduced, but which cannot be deduced from each other.* From this standpoint, therefore, the primary task of factor-analysis is to supply a systematic hierarchy of independent concepts—what I have called a 'psychographic scheme'—in terms of which any individual—whether a clinic case or a vocational case, whether delinquent, backward, or neurotic, normal, subnormal, or supernormal—may be described.[1]

Not one person, but all persons may be described in this way. As a result we discover not one child, but a number who are mentally defective, a number who are of a verbal type, a number who are introverts, and so on. The cumulative result, and indeed the ultimate aim, is thus a *classification* of these children into significant groups and subgroups—a classification that would be applicable to the entire population.

But the systematic scheme that we have worked out in this way does not merely furnish a classification of persons; at the same time it involves a classification of traits. For the psychologist is interested not only in describing individuals; he also seeks to describe their modes of behaviour. And for this purpose *the same concepts or 'factors' will be employed.* Thus, when we describe fear or disgust as an emotional rather than cognitive tendency, as having a

[1] A tentative list was drawn up in my Presidential Address to the Psychological Section at the Liverpool meeting of the British Association ([46], pp. 215–39) and expanded in *The Measurement of Mental Capacities* (Oliver & Boyd, 1927). Minor modifications have been introduced in later versions: cf. *The Young Delinquent*, pp. 22 f.; *The Backward Child*, pp. 5 f., 630 f. It should be added that this scheme serves for the examination, not only of subnormal, but also of normal cases—e.g. for purposes of vocational guidance: cf. [53], 'Schedule for Investigating Individual Cases,' pp. 10–11.

repressive rather than an aggressive character, and an unpleasant rather than a pleasant feeling-tone, we are not merely defining those emotions: we are classifying them—classifying them according to a hierarchy of subordinated and dichotomous principles (cf. p. 19). What distinguishes factor-analysis, therefore, from other ways of discovering how individuals and their numerous attributes can best be classified is chiefly this: whereas the ancient logician reached his definitions by examining the meanings of words, the modern factorist reaches his classifications by examining the correlations between the forms of behaviour to which those words very loosely refer. But the ulterior object is still the same; and, whether we are describing persons or traits, the factorial concepts adopted are simply *principles of classification.*

There is, however, a second peculiarity of factor-analysis: but this, from the logical standpoint, is only incidental. Factor-analysis, as we have seen, is quantitative. One of the first problems of any science as we understand it to-day is to devise a means for converting qualitative specifications into quantitative. The process by which the psychologist achieves this translation will more easily be understood if we think once again of the analogous but more concrete problems that arise in navigation or geography: there, too, the numerical determinations of the expert have arisen simply by progressive refinement of the cruder qualitative classifications of the plain man.

From the time of Aristotle to that of Bacon, primitive science was content with the simple classificatory procedure, such as the traditional logic continued to employ. The transition from classification in terms of qualitative attributes to measurement in terms of quantitative variables was mainly accomplished by physical science during the sixteenth century. Mental science, we are told, has still to complete the change "from the Aristotelian to the Galilean viewpoint."[1] But even in dealing with the physical world the plain man rarely thinks in terms of measurement.

[1] K. Lewin, *Dynamic Theory of Personality*. By the "Galilean viewpoint" seems to be understood the standpoint of experimental and mechanical science. The mechanical and dynamical interpretation, however, would appear to have been a secondary consequence of the more accurate ideas of

Like the navigator of old, the simple tourist does not picture his ship or his destination as something to be precisely localized on a three-dimensional diagram by the aid of orthogonal or spherical co-ordinates, with every direction and distance expressed by its own appropriate figure. He vaguely thinks of places as divided into 'hot' and 'cold,' or, if he tries spatial terms, into 'northerly' and 'southerly.' Similarly, the psychology of the man in the street is still content with a bipolar classification based on two contrasted qualities: he still divides his fellow-men into the good and the bad, the wise and the foolish, optimists and pessimists, and the like.

The primitive scientist starts from some such twofold division. Since in practice the opposite classes nearly always merge insensibly into one another, he suggests laying down a conventional line of demarcation, somewhere about the middle. Geographically, for example, he arbitrarily divides the world into what is north and what is south of the equator. When a finer classification is desired, he goes on to distinguish arctic, north temperate, torrid, south temperate, and antarctic zones—a step parallel to that taken by the Arts examiner when he elaborates the teacher's rough dichotomy of 'bright' and 'dull' pupils into terms of a fivefold scale—A, B, C, D, E. And now this manifold classification can be made finer and finer, and is ultimately expressed in terms of a specified number of equal steps—180 'degrees,' for example; presently minutes and seconds are recognized, not as sub-classes, but as fractions of a 'degree'; and thus broad classifications by qualitative attributes are completely transformed into a compact, continuous, one-dimensional scale. In much the same way the modern psychologist refines his means of discrimination, until he too is measuring intelligence on a continuous linear scale of 'mental ages' or 'I.Q.'s.' At the same time, if we are to measure with precision, we must define with precision: East and West must be clearly distinguished from North and South. Similarly, the psychologist comes to distinguish variations in sheer intellectual ability from variations in industry or laziness, the cognitive factor from the conative—two more or less independent directions which the parent and teacher very commonly confuse.

In the biological or social sciences the interaction of the deter-

measurement, which in the earlier work related to celestial rather than terrestrial movement and to astronomical rather than mechanical science. For terrestrial purposes it was virtually established when portable clocks and the Copernican theory together supplied the navigator with a means of measuring longitude as well as latitude; and thus measurement in terms of two independent 'factors' took the place of dead-reckoning.

mining variables is far more complex than in astronomy or physics, if only because of their greater number. But with the development of statistical method during the last half century, formulæ have become available whereby at almost every stage we can calculate suitable constants; and, it may be added, whether we are dealing with what the statistician would call the 'theory of attributes' or with the 'theory of variables'—with qualitative classes or with quantitative grades—his equations and his demonstrations rest at bottom on much the same principles.

Now, when we are analysing not qualitative attributes but the quantitative variables that replace them, the factorial components, as we have seen, take the form of vectors; that is, they can be represented by a line which in its turn is specified by a length and an angle; or, in other words, they vary (i) by a definite amount, and (ii) in a definite direction. Omit the quantitative element—namely, the 'amount'—and we are left with the qualitative classification according to 'direction.' Once again, therefore, we see that our factors still remain in their essential nature principles of classification, though rendered more discriminative and exact by being cast into quantitative form. Where the plain man classifies children into 'good' and 'bad,' 'bright' and 'dull,' the administrator has to lay down sharper lines of demarcation, and talks of 'guilty' and 'not guilty,' 'certifiably defective' and 'capable of profiting by a secondary education,' and the like; and the theoretical psychologist tries to improve the classification still further by thinking of the population as varying in certain directions, each more or less independent of the other and each carrying with it a number of inferable traits—as varying in what he calls 'innate general intelligence,' 'innate specific abilities,' 'innate general emotionality,' 'acquired moral character,' and the like—and then seeks to graduate these variations on a linear scale. Thus, what appear to be the measurements of a variable are merely minute subdivisions by manifold classification as distinct from the first broad differentiation reached by twofold classification; and the factorization of such variables is simply a mode of reclassifying the varying individuals according to a more refined logical procedure which has

been sharpened up until it takes the shape of an algebraic calculus.

The student, who is already familiar with the more technical procedures, can easily confirm this description if he considers the nature of what corresponds to 'factors' in modern physics. The additive 'hierarchies'[1] of the psychologist, i.e. the sets of figures that distinguish and describe the psychologist's factors, would in quantum physics be called 'selective operators.'[2] When the psychologist takes a mark list giving the measurements of, say, a hundred persons for a dozen or more correlated tests, and reduces these measurements to terms of less than half a dozen uncorrelated factors, he is following almost exactly the same procedure as the physicist who measures the magnetic moments of a number of different metallic atoms, deduces a 'spectral set of selective operators,' and so obtains the quantitative 'spectrum' of his 'assembly.' The physicist sorts his atoms into so many different kinds. In the same way the psychologist's factors are devices for taking a heterogeneous group of observables, and dividing the whole set up into smaller classes which are more or less homogeneous. Thus, what the physicist says of his 'operators' might be applied forthwith to the psychologist's factors: they provide "a species of spectral analysis in which a given inhomogeneous aggregate *a* is resolved into a number of parts which are (relatively) homogeneous with respect to some variable *y*."[3] In plain language, then, they are merely classificatory devices.

This standpoint may enable us to resolve in advance many of the other heated questions that have long divided factorists in psychology; for it now becomes evident that these controversies, though they wear a psychological or a

[1] The reader who is not yet acquainted with the peculiar use of this term in factor-analysis may turn to Appendix I, Tables I and II, for illustrations, and to p. 149 below for the definition and formula. This special usage is only accidentally connected with the more familiar meaning adopted a page or two back.

[2] See chap. IX, p. 264, and Appendix II, p. 491.

[3] G. Temple, *The General Principles of Quantum Theory*, 1934, p. 33. Elsewhere I have offered a formal proof of the virtual identity of the physicist's analysis and the psychologist's, *Psychometrika*, III, 1938, pp. 151–68.

mathematical dress, are nothing more than a resuscitation of the ancient logical and metaphysical disputes, which arose out of the problem of scientific classification, and which, at any rate for the case of qualitative attributes, have already been fairly thoroughly thrashed out.

How Many Kinds of Factors are There?—Let us begin with what would seem to be the most fundamental question of all—the issue that has been the main source of dissension—namely, how many kinds of factors are we to postulate? Spearman based his original analysis of cognition on a 'theory of two factors'—i.e. two *kinds* of factors, one general factor and a number of specific factors. Some of his followers have reduced this to a theory of a single general factor only. On the other hand, many of his critics have contended that the mind and its workings cannot be adequately described unless we invoke three kinds of factor—general, group, and specific. Others again have argued that the only set of factors that can have a real 'psychological meaning' must consist in an indefinite number of common factors, each conforming to certain limiting requirements, such as those of Thurstone's 'simple structure.'

Now, if the position maintained in the preceding pages be accepted, it seems clear that, in the first instance, the issues thus raised are points to be decided, not by considering the general nature of the mind, but rather by examining a little more closely the logical basis on which factor-analysis rests. If our factors are logical principles rather than psychological principles, we should be able to state, *a priori* and in advance, how many kinds we may in theory expect, though whether all the kinds will be found in this particular table or in that must depend upon the traits or tests actually selected for study, i.e. upon the experimental design.

The Four-factor Theory: (a) In Correlating Tests.—Consider any set of measurements obtained by testing a number of individuals for a number of mental capacities or traits. The factorist's problem, we have seen, is to classify the traits and the persons. From this standpoint it is evident that one or more of the following propositions can be asserted

about the traits, and (as we shall see in a moment) analogous propositions can also be asserted about the persons :

(1) *All* the traits possess a particular characteristic, g, and thus form a general, all-inclusive class ;

(2) *Some* of the traits possess a particular characteristic, p_1 (which the rest do not possess), and thus form a narrower sub-class; and again others possess the characteristic p_2 (which the rest do not possess) and thus form a second sub-class ; and so on ;

(3) *This* particular trait possesses one particular set of characteristics, u_1, which none of the others possess and which thus, as it were, forms a sub-class of one ; and similarly for each of the other traits.

Further, if we repeat our tests, we may be able to add that this particular trait possesses (*a*) one particular set of characteristics, s_1, *always*, i.e. every time we measure it (the series of repeated measurements thus forming, as it were, a sub-subclass), and (*b*) other sets of characteristics, e_1, e_2, ... *occasionally*, i.e. one set on this occasion, another on that.[1]

These four possibilities indicate four main kinds of factors. They may be conveniently designated (as I have elsewhere[2] suggested) by the labels traditionally used by logicians in classifying propositions according to their 'quantity.' They fall into two main groups, each of which may be redivided into two :

A. *Common* factors, i.e. those influencing several tests or traits. These are of two kinds, viz. :

(1) *Universal* or General factors, common to all the traits. Later we shall see that the general factors in their turn may take two different forms :

(*a*) *Positive* or One-signed factors, i.e. factors which can vary in one direction only, viz. above zero, never below, and whose saturation coefficients can therefore take positive values only. Usually only one such positive universal factor is distinguishable, which is then termed '*the* general factor' ;

[1] The student of logic may remember somewhat similar distinctions introduced by the schoolmen in their classification of properties : viz. *id quod pertinet* (i) *omni*, (ii) *non omni*, (iii) *semper*, (iv) *non semper*.

[2] *Marks of Examiners*, p. 259, and earlier writings.

(*b*) *Bipolar* or Two-signed factors, i.e. factors which can vary in opposite or antagonistic directions, and whose saturation coefficients may therefore be either positive or negative.

(2) *Particular* factors, each common to a certain group of traits only, and hence usually termed '*Group*-factors': they have sometimes been called 'special factors,' 'overlapping specific factors' ('overlapping,' because any single trait may contain more than one of them), or 'general factors of limited range'—phrases which all rather blur the real nature of the distinction.

B. *Individual* or Unique factors, i.e. those influencing one test or trait alone, viz.:

(3) *Singular* factors, each peculiar to a single trait, and usually called '*Specific* factors' (sometimes also 'individual' or 'unique' factors); when the characteristics they cover are regarded as irrelevant to the main inquiry, these factors, like the following, are frequently described as 'errors,' and then, in contrast to the following, are designated 'constant' or 'systematic errors.'

(4) *Accidental* factors, each peculiar to the particular occasion on which the particular trait was measured, and therefore sometimes called 'factors of error' or of 'unreliability' (the latter term in factor-analysis merely means inconsistency). Here the errors are the results, not of some gross and traceable bias, but of a very large number of very small causes. Hence the minor fluctuations for which they are responsible show the random distribution characteristic of 'chance,' and the factors themselves are consequently often called 'random errors' or 'chance factors.'

We may sum up these preliminary distinctions in what I have called the four-factor theorem. "The measurement of any individual for any one of a given set of traits may be regarded as a function of four kinds of components: namely, those characteristic of (i) all the traits, (ii) some of the traits, (iii) the particular trait in question whenever it is measured, (iv) the particular trait in question as measured on this particular occasion." This I regard as a fundamental logical postulate from which all factor theories must necessarily start. If for the sake of clearness we prefer to condense it into a symbolic equation (assuming with most factorists that the 'function' may be expressed as a weighted sum), we may write—

$$m_{ji} = \Sigma f_{jh} g_{hi} + \Sigma f_{jk} p_{ki} + \Sigma f_{jl} s_{li} + \Sigma f_{jn} e_{ni}$$

where m_{ji} denotes i's measurements in test j, the f's denote the weights or factor-loadings, the g's, p's, s's, and e's denote the four kinds of factor—general, group, specific, and error factors respectively, and the summations indicate that more than one factor of each kind may conceivably be present in each test.[1] Since, however, the kinds of factors are being defined solely according to the *number* of traits and occasions into which they enter, we can usually amalgamate factors of the same kind entering into the same trait or set of traits into one. Thus, as we shall see in a moment, except for the group-factors, all the summations may be dropped; mathematically, indeed, with only a single correlation table, it is impossible to distinguish more than one specific factor for each test, or more than one positive general factor for the whole table, or even error factors from specific.

The Need for Relevance in the Correlated Traits.—The foregoing definitions of the various kinds of factors reveal at once what a heavy onus is placed upon the initial selection of the traits. Evidently if the general factor were defined simply as that particular characteristic (or set of characteristics) that is common to all the traits selected for comparison, it could possess little or no stable meaning, unless those traits had already been selected according to some provisional principle. Much the same holds good of the so-called group-factors. At the same time, the definitions need not be taken to imply that *any* feature that is common to all the traits in a haphazard batch becomes *ipso facto* a general factor, or that *any* feature which is shared by some but not all of them becomes *ipso facto* a group-factor. We may call them 'general features' or 'group features,' but the term 'factor' usually suggests something more.

A factor we have described, not as any characteristic that may be observed or named, but as one which can serve as a principle of classification, i.e., as it were, a stable nuclear feature, which may be conceived as implying, in a wide variety of instances and under various changes of conditions, the numerous attributes that it synthesizes. As Aristotle pointed out, there are three things we must look to in seeking principles for systematic classification: the terms

[1] *Marks of Examiners*, p. 258, eq. (v).

employed at each successive level " (i) must be of the essence of the things classified ; (ii) must be taken in their right order ; and (iii) none must be omitted." [1]

(i) If, therefore, the factors are to have a stable value and a useful meaning, there must be some essential connexion between the general factor and the several traits into which it enters ; and, further, there should be, if possible, some intelligible connexion between the general factor characterizing the whole sample, and those factors that characterize the groups into which the sample is subsequently divided : so that the group-factors emerge as intelligible specifications of the general factor. If we believe in fixed genera or species dividing the natural world into a hierarchy of types, or in ' general abilities ' and ' special abilities ' as determinable characteristics possessed by the human mind wherever it is found, these factors and the traits they comprise should enable us to give an appropriate definition of those types and those abilities. If we doubt the objective existence of ' natural kinds ' and of ' real abilities,' the principles of classification ought still to have a relevance to the general purposes of the scientist's investigations. As Kelley has rightly observed, the original measurements which we analyse into components or factors must be chosen on some other basis than mere chance or mere convenience, or the simple fact that practicable tests or instruments are available for measuring them ([85], p. 66). But this relevance can be secured only by the deliberate efforts of the investigator, not by the automatic operation of a statistical technique. If the collection of traits is meaningless as a collection,[2]

[1] *Anal. Post.*, II, xiii, 97a, 23 *et seq.*

[2] A random sample of the complete population of traits I should not call meaningless. But a random sample of all existing tests (proposed by Hotelling, [79], p. 504) would have but little meaning, because the existence of a test is due as much to its practicability as to its social or psychological value, and the former has no relevance to the latter. The reader will object that we are moving in a circle. But that objection has been urged against every attempt to define what is essential or relevant : no scientific observer can ever gather a limited collection of data without tacitly making some provisional classification. On this point I must be content to refer to the more detailed analysis of the logical difficulties given by logicians themselves : e.g. H. W. B. Joseph, *Introduction to Logic*, pp. 94 *et seq.*

no subsequent 'rotating of the axes' will guarantee the emergence of 'psychological meaning.'

(ii) With this proviso, the order as well as the pregnancy of the several principles will be ensured, if we extract our factors in such a way that each accounts for a maximum amount of the variance available at each stage, and then arrange them in order of their contributions to the variance. The principle is perhaps most familiar from the parlour game of 'Animal, Vegetable, or Mineral,' where an unknown object has to be identified after a minimum of guesses: the guesser has to proceed by dichotomous questions—'Is it alive or dead?' 'plant or animal?' . . . —and the expert so chooses his alternatives, that at each successive step he eliminates the largest number of erroneous possibilities. In serious scientific attempts at classification or diagnosis, precisely the same procedure—the principle of 'progressive delimitation,' as I have called it ([53], p. 81) —is more or less explicitly adopted.

(iii) The completeness of the list will be guaranteed if the items selected for analysis consist either of the complete population of items with which we are concerned (impossible when the items are traits and hardly practicable when they are persons), or else a random, i.e. an unbiased and representative sample of that population, and if we then extract, for n correlated traits, n uncorrelated factors (if our data are free from error), or (if errors are inevitable) as many factors as the margin of error permits.

The factorization of a number of traits into the same number of factors has frequently been criticized. Thurstone lays down the 'postulate' that "the number of reference abilities in a test-battery must be less than the number of tests," and regards any other mode of factorization as uneconomical and useless.[1] The

[1] [84], p. 75. He argues that "the solution in which $r = n$ violates the fundamental postulate of science that every valid hypothesis should be over-determined by the data. The only allowable case is that in which $r < n$." To secure this he requires that r, the number of common factors, shall be identical with the minimal rank of R, the correlation matrix.

Actually the total number of factors in Thurstone's initial 'factorial matrix' is not less than the number of tests, but more, since he, like Spearman, begins by assuming that each test will have its own 'specific factor' in addition to the r common factors (p. 59). I agree that in practice specific

same objection has been advanced against my own methods of factor-analysis by Prof. Reyburn and Dr. J. G. Taylor.[1] The criticism overlooks the fact that, even when the factors do not differ from the traits in number, nevertheless they differ from them greatly in their mutual relations and in their comprehensiveness: first of all, whereas the traits are mutually correlated, the factors are (or should be) mutually independent; secondly, whereas each trait accounts for only a small fraction of the total variance, and that much the same fraction, the factors are so determined that the first or 'highest common factor'[2] shall account for as much of the

factors in the absolute sense may, like the error-factors, be conveniently excluded from the final factorial classification (though my reasons are not the same as Thurstone's). But Thurstone's postulate, while treating specific factors as negligible, makes them at the same time responsible for the greatest amount of variance that they can possibly contain.

Where we are dealing with a table of correlations, as distinct from a table of covariances, my 'summation' method (simple or weighted), if carried out automatically, does, as a matter of fact, yield factors whose number is identical with the number specifying the lowest rank obtainable for the table of intercorrelations. But that is not because I hold that a postulate of economy requires the smallest number of factors, but because that procedure happens to give definite and plausible figures for the variances. In point of fact, Thurstone's own methods would seem to involve the determination of far more common factors than I should ever venture to extract. Thus, in his illustrative analysis (*Primary Mental Abilities*) he extracts as many as 12 factors, in spite of the high probable errors, where I should have thought barely one-quarter of that number were statistically significant.

On page 92 [84] he declares himself ready to accept any method of factorizing the matrix of intercorrelations "provided the minimum rank of R is not altered"; at first sight this seems to mean that with 15 tests he would extract 10 factors (cf. Table 2, p. 77). Yet, although here and elsewhere he insists so strongly on the principle of minimal rank, he does not show how his own method will secure precisely this number of factors; and his working procedure, where "the diagonal entry may be given any value between zero and unity" (p. 108) and the computer is advised in practice to give it the value of the highest correlation, seems a double contradiction of the demand for minimum rank, since this is equivalent to demanding a minimum sum for the communalities. Indeed, the device of fitting the leading diagonal with the largest correlation from each column must almost inevitably increase the number of factors and raise it artificially well above the number indicated by the minimal rank.

[1] 'Factorial Analysis and School Subjects: A Criticism,' by H. A. Reyburn and J. G. Taylor, *Brit. J. Educ. Psychol.* (in the press).

[2] The phrase 'highest common factor' was used in my earlier articles (e.g. *Brit. J. Psych.*, III, 1909, p. 166) to denote the factor which would account

total variance as possible, the next for as much as possible of the remaining variance, and so on, and their purpose is not fulfilled unless all the significant variance is ultimately accounted for.

These two changes introduce both order and comprehensiveness into the description of the data ; and it is these that are the essential merits of a sound scientific classification, not the fact that the classes reached shall be necessarily fewer than the specimens to be classified. Select a representative sample of all the plants in Kew Gardens ; set a botanist to classify them : he begins with a comprehensive specification that will distinguish plants from other living things—the general factor ; he then divides the whole lot into seed-bearing and non-seed-bearing—a division which, as we have seen, is chosen because it carries with it larger and more numerous differences than any other ; then he redivides the former into flowering and non-flowering, for precisely the same reason ; the flowering in their turn into two contrasted classes, the one-seed-leaved and the two-seed-leaved ; and these again into orders, and the orders into varieties. Is it an objection that the varieties he ultimately names may be as numerous as the specimens he has been given ? Give him yet another plant, different from all the rest ; and he will invent yet another pigeon-hole to receive it.[1] *Economy is achieved, not by*

for a given set of correlations in such a way that the residual or partial correlations remaining would be as small as possible. This phrase seems to convey the notion best to those familiar only with elementary algebra. In more technical discussions it would seem better to substitute the term ' dominant factor,' which is used to mean the factor corresponding to the highest latent root of the correlation or covariance matrix, i.e. to the leading element in the diagonal matrix of factor-variances. This is in keeping with customary terminology in matrix algebra. In the latter the phrase ' highest common factor ' would probably suggest the leading element in the diagonal matrix obtained in Smith's reduction of a lambda-matrix to a canonical form (a related and suggestive conception, but not quite identical).

[1] Or, to use the technical language of the logician, I should argue that *infimæ species* are still *species*, and therefore in theory possess each its own factor. When we come to correlate persons, however, the problem takes a more disputable form. If each person is to have his own specific factor, does that factor represent his *principium individuationis*, or his *proprium* (his necessary properties), or merely his ' inseparable accidents ' ? (see below, p. 111). In practice the question will rarely arise. Usually there are fewer traits than persons ; and in such cases, in virtue of the lowered rank of the matrix, there will not be a specific factor for each person. When, on the other hand, there are fewer persons than traits, the persons will usually be chosen for their representative character ; and hence, once again, the person's ' specific factor ' will not really be peculiar to him individually ; it will be peculiar to him only as representing a particular class or type.

minimizing the number of factors, but by maximizing the amount of variation that each in turn will account for. If, therefore, the psychologist has a dozen tests for a dozen different abilities, which differ in a way that is at once significant and relevant to his problem, he must be prepared to admit a dozen different factors.

In practice, of course, he will probably be unable to show that they are all genuinely significant, and will usually confine his calculations to three or four factors at the most. But he must not start by limiting the amount of discoverable variance on purely *a priori* grounds at the very outset. If a solid object has three dimensions, it is not a 'fundamental postulate of science' that the axes of reference should be reduced to two; the dictates of economy are satisfied if the axes chosen are independent. Similarly, if a correlation matrix is of order $n \times n$, we must not exploit our ignorance of the diagonal entries to insist on reducing it to a rank[1] of $r = n - \sqrt{2n}$. Indeed, if the correlation matrix were a covariance matrix, such a reduction would in general prove impossible.

The foregoing requirements will become clearer still, if we turn for a moment from a study of the resemblances between traits to a study of the resemblances between persons.

The Four-factor Theory: (b) in Correlating Persons.— Most psychologists have started by correlating traits or tests. But is there any reason why we should not start by correlating persons? With this approach coefficients of correlation can be calculated which will serve to measure resemblances, not between traits, but between persons. Factor-analysis can be applied on the same lines as before; and factors of the same four types will presumably emerge.

The student who is not yet familiar with factorial research finds it actually easier (that at least has been my own

[1] With this approximate formula (accurate enough for most purposes), if the value is not integral, we take the integer next above the value given. The exact formula for determining the minimum rank is $r = n - \frac{1}{2}\sqrt{8n+1} + \frac{1}{2}$, where, if the value is not integral, we take the integer next below. With the simple summation method, the saturation coefficients are virtually deviations about an average. After the first, each column of coefficients must add up to zero; after the second column, each section of a column (cf. p. 466 and Tables VIIA ii, XIA). Thus the number of degrees of freedom is one less with each factor, i.e. n, $n-1$, $n-2$, ..., until all the $\frac{1}{2}n(n-1)$ degrees are used up. (Weighted summation has the same effect.) On summing and solving for r, the above formula follows at once.

impression) to follow the various factor theories, if they are presented to him first with reference to the classification of persons rather than to the ' analysis ' (or, as I should prefer to say, the classification) of tests or traits. Moreover, with this alternative approach, many of the foregoing difficulties disappear, or at least are simplified, because the selection of a representative set of persons seems a much simpler matter than the selection of a representative set of tests.[1]

The relation between the two approaches may be a little clearer if we compare the essential logical nature of each. When we factorize the correlations between the traits tested, we are in effect analysing the *connotation* or ' intension ' of the class selected for testing ; when we factorize the correlations between the persons, we are in effect analysing the *denotation* or ' extension ' of that class. But in either case, when we use one and the same set of observed assessments to classify both traits and persons, it seems obvious that, if our classifications are to be consistent, the same factors must be preserved, the same *fundamenta divisionis* retained, whether we start by comparing and classifying the traits or by comparing and classifying the persons.

In each case the result of the analysis will be simply to substitute a better defined and therefore more economical set of attributes for the more numerous, more detailed, and more casual set of attributes, superficially observable and more easily measured, with which we set out. And, when we turn from traits to persons, and proceed to study the resemblances between individuals and to classify those individuals according to the more orderly and pregnant attributes ultimately adopted, it is not difficult to see that, once again, our predications will be of four different kinds. Nor is it surprising to discover that the four kinds of factor empirically recognized by psychologists turn out to correspond quite closely with the four or five headings of the traditional scheme of predicables recognized by scholastic logicians—*genus*, *species* (or *differentiæ*), *proprium*, and *accidens*.

[1] The experienced statistician will see that the problem of selecting a fair sample of *tests* now gives rise to fresh theoretical difficulties ; but these are not likely to puzzle the beginner.

The Fourfold Scheme of Factors and the Fourfold Scheme of Predicables.—Let us imagine, by way of illustration,[1] a statistical analysis of the body-measurements of a mixed group of adults. The results of physical measurement will be simpler to visualize than those obtained from mental testing; and, since the study of temperament has recently led the psychologist to a renewed interest in physical types,

[1] I have here drawn the parallel between 'kinds of factors' and 'heads of predicables' only for those factors that are obtained by correlating persons. It would be equally enlightening to the research-worker (though perhaps not quite so clear to the ordinary student) to attempt the same comparison for those factors that are obtained by correlating traits. It is far easier to explain the classification of predicables by taking those predicables that can be predicated of a concrete individual subject (such as a person) than by taking those that are predicated of an abstract term or a concrete general term (such as a trait or a type). This, however, means that I have necessarily followed the later and more familiar scheme derived from Porphyry (which was concerned with the former case) rather than the original scheme set forth by Aristotle (which was confined to the latter case).

Even so, some scholars may question the interpretation of *proprium* that I have here adopted to keep the parallel as close as possible. The dispute as to what is or is not a 'property' or *proprium* is by no means uninstructive to the factorist. It raises problems fundamental to psychology which he slurs over rather than solves. In the text I have for simplicity adopted the most literal, though not the most usual, interpretation. I may add that Aristotle (whose treatment is not quite consistent) himself in certain passages allows the term ἴδιον to mean 'a peculiarity that distinguishes an individual from others' (cf. *Topics*, ϵ, i, 128b, 16 and 129a, 3–5, and Joseph's comments, *loc. cit.*, p. 107). Porphyry perhaps would have termed those attributes summed up in a man's 'specific factor' his 'inseparable accidents' rather than his *propria*. The 'properties' would then be those attributes which were causally derived from, or necessarily deducible from (which for the empirical factorist can only mean closely correlated with) the 'essential' or defining factors, i.e. those stating his *genus* and his *species* or type.

On the Aristotelian view, *everything* we can predicate about a subject must fall under one or other of the heads of predicables. It might therefore be argued that, instead of identifying predicables with factors only, they ought also to cover all the traits. Those traits that have saturations for a factor would then be 'properties' deducible from the 'essence' which that factor specified; and it is interesting to note that Aristotle himself recognized correlation (or concomitant variation) as a method of proving the relation (τόπος ἐκ τοῦ μᾶλλον καὶ ἧττον. *Top.* ϵ. viii). The Aristotelian view, however, would to my mind make too sharp a separation between the factor and the various traits it comprises. Although for simplicity I spoke at the outset of n traits, m_1, m_2, . . . m_n, as possessing a common characteristic, g, as though g were an $(n + 1)$th trait existing over and above the others (like an 'ability'

the illustration is not so remote as might be thought.[1] Having measured each person's height, and the length, breadth, thickness, and circumference of his limbs, head, trunk, etc., we can express the observable resemblances between the physical shape of the different individuals by correlating the figures by persons. The object of a factor-analysis will then be to discover the most comprehensive principles by which we may classify and sub-classify the sample population thus surveyed. Eventually we shall obtain factors of the four kinds enumerated above, each

entering into them), I shall later show that the factor is the whole pattern or system of traits rather than itself a further trait, added to, or underlying, the rest.

As for the distinction of a 'property' in the strict Aristotelian sense, it is now usually held that "in regard to organic kinds, the problem of distinguishing between 'essence' (i.e. the defining attributes) and 'property' (i.e. attributes co-extensive with the kind but not included in the definition) is insoluble" (Joseph, *Introduction to Logic*, p. 102). With this, it will be seen, I partly agree: or rather I would say that any solution is a matter of convenience or convention. As in geometry, so in psychology, we might take the 'definition' to be those essential characteristics which would enable us either to *construct* the thing defined (as in dealing with tests) or to *select* a typical sample (as in dealing with persons). The 'properties' would then be those further characteristics which (i) from factorial research we empirically find to be highly correlated with the essential characteristics and which (ii) from other lines of research we can demonstrate (or at any rate plausibly suggest) are consequences necessarily resulting from them. The whole subject, it will be seen, is closely bound up with the practical problem: how are we to construct our tests, or to select our persons and traits, so that the most useful factors will emerge from the statistical analysis? Unless our definitions tell us *what to do* in order to observe or measure the quantities defined they are scientifically useless. For that reason I should prefer to base all definitions—of traits, of tests, of sample populations, and of factors—on the *operations* required to obtain them. This principle will come to the fore later on when, it will be seen, we proceed to define our factors in terms of *selective operators* (see below, chap. ix, § iv).

[1] I may perhaps add that it has always seemed to me essential that the demonstration of physical types, whether anthropological or temperamental, should be based on a factor-analysis of the kind described in the text. It is curious that this has never hitherto been undertaken. I have given the results of a preliminary study in a recent paper ([114], pp. 184 *et seq.*). An illustrative analysis, by all possible methods, is given in my laboratory *Notes on Factor Analysis* ('II. Physical Measurements'). It is impossible to publish the numerous tables in full. Hence they are obtainable only in roneo'd form.

factor being specified by a particular pattern of bodily measurements.

(i) The general factor responsible for the majority of the resemblances will obviously describe the general physical shape characteristic of all human beings as such—the ideal form or *Gestalt* that distinguishes man from all other creatures. This first factor therefore defines, in terms of physical measurements,[1] the persons' essential humanity : in short, it states the *genus* to which the whole sample belongs.

(ii) In certain sub-groups, however, we shall find marked and typical deviations from this generic shape. Half the individuals, for example, will have broader hips, narrower shoulders, and tinier extremities than we should otherwise expect ; the remaining half will show the reverse peculiarities. This twofold subdivision can readily be expressed by two subsidiary factors, dividing the entire sample into two sub-groups or *species*. Since the first species is the negative of the second, we can, if we prefer, reduce these two antithetical 'group-factors' to a single 'bipolar factor.' The group-factors we may regard as specifying the female *species* and the male *species* respectively : the bipolar factor we may regard as indicating a *differentia*, i.e. as differentiating the females from the non-females. No doubt, if we continued, we should encounter other group- or bipolar factors differentiating old from young, broad-headed races from narrow-headed races, Kretschmer's 'pyknic' type from the 'leptosomic,' and so on.

(iii) Since no individual is a perfect specimen of the species or type to which he belongs, but always varies slightly from its ideal or average shape, and that in a way which is unique, we shall ultimately reach a third kind of factor, one for every person, and each character-

[1] It will perhaps be the artist's view of humanity rather than the anatomist's, Sir Joshua Reynolds's '*central form* from which every deviation is deformity' rather than Quetelet's *homme moyen*. For centuries, writers on anatomy for artists have endeavoured to deduce ideal proportions of the human figure for the benefit of the art-student, and have often appended a note saying how certain types (the different sexes, the young and the old) differ from the generalized ideal : in so doing, they have, as it were, carried out a rough factor-analysis of the type described in the text. The anatomist usually gives simple and unweighted averages ; but for his famous 'canonical statuette' Carus seems to have based his proportions not on simple averages of ordinary persons, but rather on weighted averages (as we should call them), i.e. he gives more value to the measurements of a limb as found in persons in whom that limb shows a relatively perfect development (*Symbolik des Menschlichen Gestalts*, 1857).

istic of that particular individual only. Such a 'specific' factor sums up what the logicians sometimes called his *propria*.

(iv) We should, however, only regard these individual variations as genuinely characteristic of—or 'proper' to—the individual if they showed some constancy, appearing and reappearing every time he was measured. Actually, on repeating the measurements, we shall not get exactly the same set of figures; and these minor variations we shall regard as 'random fluctuations' or 'errors of measurement'—in a word, as *accidents*.

General factor, group-factors (or bipolar factors), specific factors, and chance factors—this fourfold scheme is thus in effect an independent rediscovery, in a special application, of the fourfold scheme of predicables handed down by traditional logic. Indeed, nearly every one of the fundamental controversies that have perplexed the psychological factorist could be paralleled by the ancient disputes that have arisen out of the famous schemes originally set forth by Aristotle and his commentator Porphyry.

Why, then, the mathematical disguise? Simply because, as we have already seen, the popular notion of personal characteristics as being merely present or absent, and of individual persons as being assignable to a few non-overlapping classes, each comprising relatively homogeneous members, is far too crude and inexact for a scientific description. Hence, what were originally plain principles of qualitative classification have taken on a quantitative form, and have become 'factors' instead of mere *fundamenta divisionis*. The observable traits in terms of which they have to be characterized cannot be accurately specified without introducing some kind of measurement. Such traits cannot be abruptly divided into the essential and the inessential, those inseparably present and those invariably absent. They vary, in almost every case, not only discontinuously or in kind, but also continuously and in degree. In consequence, the resulting classes—certainly those that we can observe, if not those that we can infer, the *phenotypes* as distinct from the *genotypes*—cannot, as Aristotle and early logicians assumed, be treated as clear-cut groups, with no transitional forms. Hence verbal methods of classification must be supplemented and checked by graded

methods. We are forced first of all to measure the traits and persons, thus starting with 'variables' formed by the numerical measurements, instead of with mere 'attributes'[1] named in verbal terms; then to measure their resemblances, thus obtaining their covariability or correlation; and so ultimately to measure (i) first, the diagnostic importance of the traits—which we can specify by the 'saturation coefficients' for traits when correlating traits, and (ii) secondly, the closeness with which the individual approximates to a typical member of his class—which we can specify by the 'factor measurements' for persons when correlating traits, and by the saturation coefficients for persons when correlating persons. It is for this reason, therefore, that what is essentially a logical analysis masquerades as an arithmetical analysis.

The logician, however, will at once inquire whether this step is justified. No doubt, he will argue, it will be easier to 'factorize' our phenomena if we first substitute figures for facts; but is such a substitution valid? As a logician, he will remind us, he has laid down the postulates to which measurable quantities must logically conform.[2] Has anyone formally demonstrated that these postulates are satisfied by the factorist's material? The factorist, we have seen, sets out with an additive equation. Has he ever shown that mental abilities, mental performances, or mental characteristics generally are really additive, or is this a gratuitous assumption?

The Postulates of Mental Measurement.—In view of cur-

[1] As before, I use the terms 'attributes' and 'variables' as defined by Yule ([110] pp. 11, 82). A friendly critic maintains that my statement "seems to obliterate the insuperable distinction between the statistical theory of attributes and the statistical theory of variables." My reply is that Yule and Kendall have already sought to "remove any possible idea that the theory of attributes is concerned solely with qualitative classification and is not also appropriate to the more precise data given by a numerically assessable attribute" (p. 78).

[2] Cf. Conrad and Nagel, *Introduction to Logic*, chap. xv; Stebbing, *Modern Introduction to Logic*, chap. xviii, § 5; Johnson, *Logic*, Pt. II, chap. vii. The possibility of measurement in *general* psychology has, of course, been very fully discussed; here I am referring to the measurement of the mental performances or qualities of individuals.

rent controversies, it is particularly desirable not to shirk this issue. I shall, therefore, endeavour to state what this transformation from a qualitative classification to a quantitative grading seems to presuppose. "Attributes," says Stephenson, "cannot be converted into variables in the twinkling of an eye." "To calculate a correlation coefficient," says Thomson, "is to assume that the marks [correlated] are in some sense commensurable." And both writers regard the requirements of valid measurement or marking as fatal to certain proposals, recently made, for applying factor-analysis to new types of psychological material, and in particular to certain modes of correlating persons. Representatives of the 'intuitionist' school go further; they maintain that every form of factor-analysis is vitiated from the start, since the very attempt to apply quantitative measurement to 'living personalities' "implies views that are not merely false but meaningless." "No one can safely assign a figure to any mental quality or mental product."

Their arguments, couched more or less in their own words, may be summarized as follows. "Lengths, weights, and times," they say, "can be legitimately measured in scientific units; they are unidimensional variables. The factorist proposes to measure traits and persons in the same way. But the variables he has to factorize do not differ simply in numerical magnitude. They are qualities, not a set of quantities: we cannot add units of intelligence as we add twelve inches to make a foot; nor can we put samples of behaviour into a balance, and subtract and multiply them as we subtract and multiply ounces and pounds. It follows, therefore, that to take over the mathematical methods of the physical sciences and use them for the purposes of psychological description and analysis will commit us to a fundamental flaw." [1]

[1] Cf. 'Measurement *versus* Intuition in Applied Psychology,' *Character and Personality*, VI, pp. 114–31 and refs. In general psychology (as distinct from applied) the dispute over 'measurement *versus* intuition' is at least as old as Leibniz, who maintained that, since mental phenomena could not be represented as continuous variables, any mathematical treatment of psychological problems was doomed to fail: cf. Münsterberg, *Grundzüge d.*

That difficulties, and even fallacies, may be involved in applying simple arithmetic to living personalities, I readily admit. But the objections as stated rest on a very narrow notion of what mental measurement entails; and, since none of the writers explicitly enunciates the presumable requirements, I shall attempt to set them forth here.

To measure is "to estimate (an immaterial thing, person's character, etc.) by some standard or rule."[1] If 'etc.' may be held to cover anything that belongs to a definable class and possesses the necessary properties, whether or not it is 'immaterial,' then this definition will serve for measurement in physical science as well as in psychological. Every measurement calls for an experiment. Nevertheless, measuring as such even in physical science is a mental process not a physical: it consists, not in adding the weights, but in reading the scale where the pointer points. Thus, contrary to the implications of the argument just quoted, the mere circumstance that instruments have to be used for the more exact forms of measurement does not turn it into a physical process, any more than the use of a telescope turns observation into a physical process. "C'est toujours avec nos sens que nous nous servons de nos instru-

Psychologie, I, pp. 260 f. The reader may also refer to the recent *Report* of the British Association Committee on 'Quantitative Estimates of Sensory Events' (Dundee Meeting, 1939) and the admirable Presidential Address of Dr. R. J. Bartlett at the same meeting. I should add that the above was in type before I had an opportunity of reading either of these suggestive papers, and must therefore confine my comments to footnotes.

[1] *Oxford Dictionary*, s.v. The physicist adopts a slightly narrower definition. "Measurement is the process of assigning numbers to represent qualities: the object of measurement is to enable the powerful weapon of mathematical analysis to be applied to the subject-matter of a science" (Campbell, *Physics: The Elements*, p. 267). I have a mild preference for the broader definition given by the dictionary, first, because I do not hold that mathematical analysis is limited to numbers; and, secondly, because I hold that the assignment of non-numerical marks (e.g. the 'alpha minus,' 'beta plus,' etc., of the Arts examiner) is also a form of measurement. The fact that his method is inexact does not prevent it from being a form of measurement, though the fact that it is usually *very* inexact makes it usually a bad form of measurement. But "in the sense understood by the man in the street, exactness has almost disappeared from the subject-matter of modern science" (Jeffreys, *loc. cit.*, p. 214).

ments." [1] A man with good auditory imagery could construct a whole-tone scale without any apparatus at all.

Taking the notion of class as fundamental,[2] it would seem that the properties the psychometrist has to consider are of two kinds: the elements of the class must be subject (1) to certain *relations* and (2) to certain *operations*. The former chiefly determine the requirements of *intensive* magnitude; the latter those of *extensive* magnitude.

(1) *Relations.*—To construct a 'unidimensional variable' what the factorist primarily needs is not a 'set of quantities,' but merely a linear series. To transform a classification into a grading, all that is needed is to convert each of the 'classes' into an 'ordered class.' This can be done *by the aid of exclusively logical notions*, without invoking any 'arithmetical' concepts or the 'mathematical methods of the physical sciences' as ordinarily understood. And no one, I presume, will criticize the factorist for trying to be more logical.

To arrange traits, personalities, or anything else in order, it is necessary and sufficient to find a relation that is (i) connexive, (ii) asymmetrical, and (iii) transitive, and to demonstrate by empirical observation that this relation holds good of the members of the class. Thus, if x, y, and z denote

[1] H. Poincaré, *La science et l'hypothèse*, p. 36. The tendency of the materialistic physics of the nineteenth century was to assume that absolute space, absolute time, and quantity of matter were alone directly measurable, and to suppose that measurement consisted essentially in the physical division of lengths, durations, and masses into additive unit parts. More recent developments alike in physics and in mathematics would seem to render this view rather difficult to sustain; yet it appears to persist whenever the critic discusses the possibility of mental measurement.

[2] The modern notions of measurement, like those of mathematics, are ultimately derived from Cantor's theory of classes (*Mengenlehre*): *Math. Ann.*, Vol. XV–XLIX, 1872, *et seq.* In this country his views have chiefly been developed by Russell. A most important contribution, which seems far better known to mathematicians and logicians than to psychologists, is Meinong's paper *Uber die Bedeutung des Weber'schen Gesetzes*, 1896 (reprinted *Ges. Abhandlungen*, 1913). In the paragraphs above I have borrowed my postulates mainly from Russell (e.g. *Introduction to Mathematical Philosophy*, 1919, pp. 29 *et seq.*) with some slight modifications suggested by later writers (cf. N. F. Campbell, *Physics: The Elements*, 1920, and *Measurement and Calculation*, 1929; also Johnson, *Logic*, Pt. II, chap. vii).

possible members of the class, the requisite conditions may be formulated as follows :

(i) *Connexive Postulate.*—If x and y both $< z$ or both $> z$, then either $x < y$, $y < x$, or $x = y$.

(ii) *Postulate of Asymmetry.*—If $x < y$, then neither $y < x$ nor $y = x$.

(iii) *Postulate of Transitivity.*—If $x < y$ and $y < z$, then $x < z$.

Here $=$ does not necessarily mean ' equals,' but merely ' may always be interchanged in the argument ' ; and $<$ does not necessarily mean ' is less than ' but merely stands for *any* relation obeying the conditions specified (e.g. such a relation as " precedes," " is nearer than," " more difficult than," " preferable to," " commoner than," " happier than," " redder than," " more beautiful than," etc.).

Having established the existence of such a relation between all the members of our psychological class, we may employ any convenient set of symbols to represent the series thus constructed—the numerals or the letters of the alphabet in their conventional order or position along a line or down a column. All these are characterized by order as above defined ; and we can apply them by a one-one correlation to the members of any given class that can be ' linearly ordered,' without introducing the ' rules of arithmetic,' or any of the ' notions of numerical magnitude ' to which intuitionists so strongly object. For example, in the Report on *Mental and Scholastic Tests* there is a scale of 10 drawings, purporting to be arranged in an ' order of merit.' A few minutes only are needed for the experimental check ; and practically every observer will agree at once that the arrangement fulfils each of the simple requirements enumerated above. A scale so defined forms only the beginning of a scheme of mental measurement ; yet, so far as it goes, it seems every whit as valid and as useful as Mohs' well-known scale of ' hardness ' used by mineralogists (" diamond = 10, ruby = 9, . . ., talc = 1 ") or Beaufort's familiar scale for reporting the strength of winds (" calm = 0, light air = 1, . . ., whole gale = 10 ").

The construction of such a scale, however, is by no means so easy as its verification. In practice, when the psychologist

sets out to show that certain items or certain individuals obey the three postulates just stated, he finds several troublesome obstacles and entanglements appearing in his path. To begin with, mental qualities, as named and described in popular parlance, are highly complex. Consequently, when we try to classify or grade things according to mental qualities, the experiment may break down at the outset. Widely different results are reported by different observers, and even by the same observer on different occasions. Hence the hasty conclusion has often been drawn that mental phenomena by their very nature are not amenable to measurement.

For example, in laying down the principles of scientific measurement, Campbell begins by announcing that, although we can measure the number, weight, density, and possibly hardness of a group of physical objects, we cannot measure their colour or their beauty.[1] Let us consider his two examples, since they are precisely the things the factorist would like to measure. Let us take colour first, and attempt the necessary experiment to see how far they conform to our three relational postulates. All observers will agree that crimson lake (x) and gamboge (y) are both less blue than ultramarine (z). Accordingly, let us compare the first two colours in terms of the same relation. At once the difficulties begin. One observer declares that crimson lake is less blue than gamboge, because "gamboge has a touch of green in it, and therefore of blue"; another, that the gamboge is less blue than crimson lake because crimson lake "contains a little purple"; a third, that the two are "interchangeable, at least as far as blueness is concerned." The vast majority insist that the question put is meaningless. Must we then conclude with Campbell that "there is no natural order of the colours," and that "the assignment of numerals to colours is arbitrary, because it is not dictated by judgements which form part of the subject-matter of science and for which universal and impersonal assent can be obtained?"

If so, we shall hardly think it worth while to collect similar observations in regard to relative beauty. We shall be inclined to acquiesce forthwith in Campbell's view that any attempt to measure beauty would be "fantastic." "No agreement," he asserts, "can be obtained for judgements concerning it." The end of the argument is easy to foresee. Only judgements regarding length,

[1] *Loc. cit.*, p. 268.

weight, number, and perhaps a few other physical properties can form "part of the subject-matter of science," because for these judgements alone "universal and impersonal assent can be obtained" [1]—a statement which might rather surprise lay members of the British Association who had dropped in at certain meetings of Section A.

Much the same reasoning has been used by psychologists themselves when discussing the assessment of persons or of personal behaviour—for example, in their criticisms of the attempt to rank different individuals for the same mental qualities and (more often still) the more recent attempts to rank different qualities for the same individual. But with all such problems the first essential is to make the experiment. And this is precisely what the factorist proposes. He starts by denying the tacit assumption that the only alternatives are either 'no agreement' or 'universal assent.' [2] Between psychologists, and presumably between other men of science, agreement may be of varying degrees; and these degrees the factorist can easily assess by a coefficient of correlation. To his surprise he finds, with a suitably chosen set of pictures, quite as high a consensus of opinion among art critics about their relative beauty

[1] *Loc. cit.*, p. 273. Cf. Brit. Ass. Report on *Quantitative Estimates of Sensory Events*, p. 16: "physics is the discovery and study of those relations between sense-perceptions concerning which *universal agreement* can be obtained."

[2] It is not my object to criticize Dr. Campbell's admirable exposition. From his later chapters on 'errors of measurement' it is clear that the statements quoted above from his chapter on 'the first conditions of measurement' were expressed in an unqualified form solely in the interests of lucidity. Nevertheless, these statements have frequently been cited by others to support the view that measurement in psychology is impossible. In an earlier chapter still he explicitly recognizes "degrees of knowledge"; and suggests that the "degree of knowledge is measured by the subjective mental discomfort we should suffer if we found it was not true" (p. 160). When endeavouring to show the objective or impersonal character of judgements of beauty, I carried out parallel experiments on judgements of truth. In both cases, according to the introspections, 'degree of mental discomfort' was a common criterion. But, even with experts in each of the fields, the amount of agreement (measured by a so-called 'reliability coefficient') was often lower for scientific opinions than for æsthetic. If (as one reader suggests) it might be better to "base our criterion of what is, or what is not, the subject-matter of science, not on agreement between experts, but on agreement with the facts, i.e. power of prediction," then I may add that the psychologist's prediction of what a child will do in certain tests to-morrow is far more accurate than the meteorologist's predictions of what to-morrow's weather will be.

(even though the critics are drawn from very different schools) as apparently obtains among physicists in regard to the mass, distance, or speed of movement of (say) some of the remoter stars.

It is not difficult to see why our first attempt at a serial arrangement of colours broke down. Colours vary in several directions at once; and to arrange a set of mixed variables in a one-dimensional scale is obviously impossible. It is like trying to state the position of the stars by assigning a single numeral to each. If, however, we carry out a factor-analysis for colours, we shall easily be able to show that they can be ordered in a perfectly coherent system, provided three bipolar factors are employed, i.e. provided we use three independent generating relations, instead of one.

A similar confusion stultifies a good many of our first efforts at serial arrangement in applied psychology. When a teacher is asked to grade the pupils in his class, his judgments will often fail to satisfy our preliminary postulates. He puts Tom above Dick "because Tom is brighter"; and Dick above Harry "because Dick has learnt far more since he has been with us"; but presently he will decide that Harry "ought certainly to go above Tom," because "Harry is far more intelligent, although he is rather lazy." Since the postulate of transitivity is thus violated, must we infer that any order of merit for school pupils is out of the question? The answer obviously is that "merit" denotes a complex and therefore somewhat ambiguous relation. Once again a factor-analysis is needed to ascertain whether the complex relation is not analysable into two or more that are independent of each other—relative intelligence and relative industry, for example, and possibly relative speed of memorization. As soon as we have extracted a suitable relation the task of satisfying the postulates is comparatively simple.

But even when these first three postulates are satisfied, further practical difficulties arise which suggest that the three requirements alone are not enough to supply a satisfactory means of grading. First, as we have seen, a single observer and a single set of observations are seldom likely to be conclusive. If in playing chess Tom beats Dick at the first game and Dick beats Harry at the second, we cannot be sure that Harry will not beat Tom at the third. Evidently, therefore, we must repeat the trials or the tests. But that will produce differences between the several relations that are differences not of kind, but rather of degree: x beating y at 19 games out of 20 is a different relation from x beating y at only 11 out of 20.

Secondly, suppose the master has successfully ranked the 20 pupils in his form in order: what is he to do if six new boys arrive? Two perhaps may require to be placed between the former 2nd and 3rd, and three between the former 8th and 9th; and the brightest may be better than any he had before. Yet a simple order of merit, based on consecutive integral numbers, makes no allowance for possible gaps. On the other hand, if he merely renumbers them, the figures for most of them will be altered, and the last boy will be called no longer 20th but 26th.

Thirdly, the existence of such gaps, and of gaps differing in size, is a subject of constant comment from the teacher. His task is like that of the earlier chemists, who sought to rank all the elements according to their atomic numbers: at certain points they felt compelled to leave extra wide spaces, though at the time there was no known element to insert between the neighbours. In the same way, simply to number the pupils in order does not convey all the relevant knowledge that an observant teacher could infuse into his grading. Nearly always, for example, the intervals between the first boy and the second, and again between the last and the last but one, are more glaring than the tiny differences between pupils near the middle of the list, whom he is tempted to bracket as 'ties.'

It seems evident, therefore, that we require some rule whereby (i) the order of pupils can be described regardless of the number of pupils in each particular batch and (ii) the numerals used to indicate the order shall also indicate the distance between the successive members. Even if we grade by consecutive relations, we surely need cardinal numbers rather than ordinal to express the amount of difference.

The plain teacher, who knows nothing of 'percentile ranks' or fractional 'grades,' thinks first of a simple solution. Could we only assume that every mental quality—intelligence, intensity of sensation, amount of pleasure, or impression of beauty—was formed each out of separable parts coalescing into a continuous whole, as raindrops coalesce to form a pool of water, then his task would be easy. If intelligence, for example, consisted of atomic elements, and if Tom's sample contained x elements and Dick's contained y elements,

we should merely have to devise some way of counting or estimating the number possessed by each. Nor is the proposal so far-fetched as it might seem: for, even if we cannot count unitary elements of ability or emotion, we can count the number of unitary reactions; e.g. the number of words Tom spells correctly in a uniform list of 50, or the number of times he loses his temper within a specified probationary period.

Many writers, indeed, have explicitly assumed that the usual correlation formulæ and factor can only be applied upon some such atomic postulate. Critics and defenders of factorial methods alike apparently believe that the methods are illegitimate, unless the measurements to be factorized are 'extensive' measurements. Cattell and Vernon in their recent controversy, for instance, argue as if the issue turned essentially on the question, whether qualities of personality can be measured in the same way as length and weight. "Clearly," says Vernon, "our attempts at measurement, whether with tests or with ratings, inevitably disrupt the personality into separate bits, such as can be handled by our quantitative techniques; and naturally lead to the theory that personality consists of such bits." [1]

Now, I have argued above that measurement is not necessarily a physical process, and that wholes may be measured without implying that those wholes are divisible into separate parts. Even the measurements of the physicist do not always involve direct counting or direct superposition. To measure density we need not count the number of molecules; nor do we estimate temperature by putting one heated body on top of another or adding the component temperatures. Nevertheless, although our correlation formulæ do not require us to sum separable elements, they require us to sum differences. That is true even when we correlate orders or ranks; and I should agree that, unless the rank-differences can be treated as approximately equal, the ordinary formula for rank correlation is strictly inapplicable.[2]

[1] *Character and Personality*, *loc. cit. sup.*, p. 2.

[2] Thomson seems to assume that, in general, correlating persons is only valid "where each person can put the 'tests'" (i.e. items to be judged—pictures, subjects preferred, emotional traits of other persons) "in an order of preference, according to some criterion or judgment." But, when we come to subtract the rank-numbers and add the differences, we are going beyond the postulates of mere order or ranking. We are making assumptions about the spacing of the rank-numbers; and, if there is any reason to believe —even the slender reason of subjective impression—that the spacing is much wider at some parts than at others, then I should hold that the rank-formulæ are strictly speaking invalid.

Distensive Magnitude.—The dilemma can be solved by recognizing a mode of measurement which is not extensive, but is yet additive. Relations themselves may have magnitude. Thus, as Russell points out, " the difference or resemblance of two colours is a relation, and is a magnitude; for it is greater or less than other differences or resemblances." [1] We are thus led to the concept of what has been called ' distensive magnitude.' " By distensive magnitude is meant degree of difference, more particularly between distinguishable qualities ranged under the same determinable." [2]

Thus the difference, interval, or ' distance ' between the first and last in our scale of children's drawings is plainly far greater than that between the first and the second. We can symbolize the difference by writing $(10 - 1) > (2 - 1)$. And, if the initial items can be arranged to form a transitive asymmetrical series, then differences between them can also be arranged to form such a series: e.g. $(10 - 1) > (9 - 1) > \ldots > (2 - 1) > (1 - 1)$. Here $>$ now means ' perceptibly greater than ' ; and we are thus dealing directly with magnitude in the literal sense. In theory, we should be able to proceed step by step from these differences to differences of higher orders, until at last we reach equal differences, whose differences in their turn would vanish. In practice such a proceeding is scarcely feasible; and a more reasonable plan is to begin with differences that are barely perceptible, and, as it were, work backwards.

It will be observed that zero for a distensive magnitude indicates equality, e.g. $(10 - 10) = 0$, whereas zero for an intensive magnitude (with which it is currently confused) indicates non-existence. This distinction will save many common fallacies. To distinguish distensive from extensive magnitude is still more important. The interval between the notes C and E is not formed by literally adding more notes to the interval between C and D; and the interval of ' a third ' is not really measured by the three notes that it comprises, but by the two whole tones (a tone being not a note, but an

[1] *Principles of Mathematics*, p. 171. This view is at least as old as Leibniz. " As for the objection that, although space and time are quantities, order and position are not, I answer: order also has its quantity. Relative things have their quantity as well as absolute things. There is distance or interval." (*Philosoph. Werke*, VII, p. 404).

[2] Johnson, *loc. cit.*, p. 169.

interval) which separate the extremes and which the three notes mark off.[1]

If the formulæ of correlational and factorial analysis are to be applied to such magnitudes, it will be necessary to show that the differences, distances, or intervals are subject to the operation of addition. This holds in attempting to measure both the factors themselves and the empirical variables from which they are derived. The possibility of finding a unit of measurement follows from this, not *vice versa*, as Stephenson's criticisms seem to imply.

First of all, however, let us see what procedures have actually been employed in attempting to replace crude ordinal results (which are all that the teachers' usual method of marking really supplies) by a set of additive measurements. The construction of mental and scholastic tests exemplifies the expedients most commonly adopted. For the measurements in individual psychology four main types of scale may be distinguished: the student will find concrete illustrations of each in the L.C.C. *Report* already cited [41].

(i) *Speed* or frequency scales. For these the initial measure is generally either (*a*) the time taken for a specified number of reactions (amount-limit method) or (*b*) the number of reactions performed in a stated time (time-limit method)—the stated time being either a short period measured in minutes, or a longer period measured in

[1] The distinction is clearly drawn by Meinong (*Uber die Bedeutung des Weber'schen Gesetzes*, p. 22), who distinguishes the *Unterschied* from the *Verschiedenheit*, e.g. the *Strecke* (or 'stretch') between two points from the 'distance' between them: the former is itself a line; the latter is not. The psychologist will recollect Delboeuf's proposal to measure sensory intensity in terms of 'sense-distance' (*contraste sensible*), instead of in terms of the sensations themselves as Fechner had proposed (*Revue philosophique*, 1878, V, pp. 53 f.). What I may call Fechner's fallacy constantly reappears in criticisms of psychological measurement; e.g. in the British Association *Report* the "measurability of sensation intensity" is denied because a sensation "cannot be analysed into parts from which it can be resynthetized by a defined operation of addition" (*loc. cit.*, App. IV, p. 15). Nevertheless, as the instance cited in the text suggests, it may be adequate for practicable purposes to estimate distance by the corresponding 'stretch,' the abstract interval by the number of concrete things we can interpolate within it.

years, as in estimating development or achievements at different ages (e.g. [41], pp. 273, 300).

(ii) *Graded* scales. In these, instead of being all of equal difficulty, the test-items are arranged in order of increasing difficulty, proceeding by approximately equal steps. The comparative difficulty of each test-item is usually determined by the proportion of examinees giving correct responses (e.g. [41], p. 138, etc.). With this and the preceding type of test the initial figures will not necessarily space out the positions on the scale at equal intervals. In psychology, as in physics, the requirements of equality may oblige us to take, not the experimental figures themselves, but some function of them that will yield more plausible results—that will lead, for example, to laws or correlations expressible in linear form.

(iii) *Impressionistic* methods: qualitative scales with subjectively determined units. In these the equality of units is decided introspectively. Consciously or unconsciously, the observer relies on his judgment of just-perceptible differences or of equal-appearing intervals, much as he does in estimating time or length when no subdivisions on the dial or measuring-rod are visible. For more accurate results the observers and the observations will be multiplied, and the figures adjusted by statistical treatment—e.g. by averaging, or by invoking the familiar principle that 'differences noticed equally often are equal.'

(iv) *Analytic* methods: qualitative scales with objectively determined units. The total product is analysed into a number of separable or distinguishable elements (for example, the component test-processes in the booklets for testing intelligence, or what are supposed to be the essential elements of good writing in tests of English composition). The mark for the whole is then obtained as a weighted or unweighted sum of the several parts. Except in the simplest cases, this in itself really entails a preliminary factor-analysis; but, unless a definite experiment is first carried out (such as that described below), there is no guarantee that the intervals denoted by the same figure in different parts of the scale will really be equal (cf. [41], pp. 308, 331; and [134]).

(2) *Operations.*—Both the construction of such scales, and their application to new examinees, call for some principle whereby we may determine equality, i.e. some clearly defined operation of 'matching' by means of which we can decide whether the fundamental differences (or 'intervals') are themselves different or not. What is meant by saying that two such distances are interchangeable? Physical

interchange is impossible; we cannot move the intervals about as we transport the interval on a foot-rule from one object to another. Theoretically, we may define equality as the limit that is reached when a difference is reduced indefinitely (e.g. when in tuning a violin we diminish the intervals between A♭ and A or between A♯ and A as far as we can). Practically, we shall assume $x = y$, when we can no longer decide that $x > y$ or that $y > x$. Such a decision should be based, not on one observation, but on many, and will lead to statistically defined criteria, like those adopted in the ordinary psychophysical methods.[1]

From a formal standpoint the central problem will be to show that the intervals or 'distances' form a 'group' for the operation of addition. If possible, therefore, we have to demonstrate, not only that they form an asymmetrical and transitive series, but also that they conform to the further postulates that addition logically presupposes. *Such a demonstration can only be carried out by actual experiment.* Hence the result can only be that the postulates hold *approximately*, i.e. after due allowance has been made for a certain amount of inevitable error. The fact that an experiment is necessary, and as such will generally involve a physical operation, does not mean that the process

[1] Some of the more obvious devices, mainly based on the traditional psychophysical methods, were described in my earlier reports; they are more fully and systematically set out in such works as Guilford's *Psychometric Methods*. The conversions proposed themselves rest on additional assumptions for which both the empirical evidence and the *a priori* arguments are often far from convincing, e.g. that time-measures form a geometrical rather than arithmetical scale, or that frequencies are distributed in correspondence with the normal curve. My own view is that in theory we should start with a more general form of conversion (logistic rather than logarithmic in the first case, hypergeometric rather than normal in the second), and seek experimental data for the requisite constants. In practice an empirical procedure is usually sufficient, if checked by the results obtained with other types of scale. Let me add that in the early controversy between Prof. Karl Pearson and myself over the non-linear character of the Binet age-scale [27], I did not mean to imply that such scales could *never* be made linear (or sufficiently linear for practical purposes), but merely that the assumption of linearity in the original unstandardized version required preliminary testing and (probably) considerable readjustment within the scale itself (cf. [41], p. 139, 'diagrammatic representation of the test-series as a linear scale').

of addition itself must be physical, as those who are pre occupied with physical measurement are prone to suppose.[1]

The general procedure may be illustrated by a tentative study carried out at my suggestion by W. E. Craven on the measurability of children's handwriting.[2] A specimen script was obtained from

[1] If a draughtsman wants to determine whether one side of an oblong is twice the length of the other, he *may* find it easier to manipulate the paper, by folding or the like, so as to get one length exactly beneath the other. Similarly, in comparing ' distances ' between two pairs of pictures, the person judging usually likes to bring the two pairs side by side. But neither action makes the comparison itself a physical operation.

It is only fair to observe that the view I am criticizing is by no means confined to physicists. Thus Conrad and Nagel (*Introduction to Logic*, 1934, p. 297) explicitly declare that, unless we can demonstrate a " *physical operation* of addition " (their italics) corresponding to the process of arithmetical addition, we cannot treat the measurements obtained as additive. " It is nonsense," they maintain, " to say that one person has twice the intelligence of another, because no operation has been found for adding intelligence " : any such statement would be " strictly without meaning." " When we assert that one man has an I.Q. of 150 and another one of 75, all we can mean is that in a specific scale of performance . . . one man stands ' higher ' than the other " (pp. 294, 298).

Now, if it could be shown that performance in the intelligence tests could be ascribed to the same general factor at every age, and that the curve of growth in that factor was linear, then I should reply that it was by no means " without meaning " to say that a normal child with a mental age of 10 had twice as much ' intelligence ' as one with a mental age of 5. More generally, it would seem quite permissible to say that one person had twice the capacity for mental work as another. However, with this mode of multiplication the correlationist is not concerned. On an I.Q. scale the true zero is not an I.Q. of zero, as the criticism cited presupposes, but an I.Q. of 100. The correlationist when he multiplies (for weighting and the like) works always with differences or with ' deviations,' as he calls them, not with absolute figures. The form of statement whose validity he has to vindicate is not $2 \times 75 = 150$, but $2 \times (125 - 100) = (150 - 100)$. Whether he starts by calculating correlations or covariances, his initial measurements are first expressed in a ' distensive ' form : he takes, not the figures that specify the observed positions, but the distances or intervals between them.

[2] The general procedure was based upon a rough set of experiments attempted by Miss Pelling and myself when endeavouring to select a series of pictures for testing pictorial preferences. Our object was to test the postulates of addition and linearity, not for a sample of the entire ' field,' but only for specimens selected to form a would-be linear scale. We found that with widely spaced scales (like the age-scales in the L.C.C. *Report*) the postulates were adequately satisfied ; but, for testing preferences, the coarser

every child in a given school (157 in all). The method of 'incomplete paired comparison' was followed. The specimens were first roughly grouped and ranked; and more or less similar pairs were then submitted to 63 students and teachers. They were first asked: is this specimen of greater, less, or equal merit than that? Merit was defined as 'legibility, i.e. ease and comfort in reading.' If more than 40 observers gave the same answer, the judgment was held to be 'significant,' and the group was said to agree. Thus (so it was eventually claimed) the group 'agreed' that the entire[1] 'field' of 157 specimens could be arranged in an asymmetrical, transitive series, such that every specimen was either of greater, lesser, or equal merit as compared with any other; and the same for the differences. On this basis a final order was drawn up, proceeding, where possible, by 'just perceptible gradations'; and the specimens were numbered on a decimal system, 0·1, 0·2, . . . 0·9, 1·0, . . . etc. The 'sum' of two intervals, say (8·0 — 7·0) + (7·0′ — 6·0), was then defined as the interval between the extremes, viz. (8·0 — 6·0), where 7·0 and 7·0′ do not necessarily denote the same specimen, but only specimens having the same number and therefore indistinguishable in merit.

With this definition the following postulates were verified, for a significant majority of the 63 observers. For simplicity I formulate each postulate in terms of a concrete instance.[2]

(i) *Postulate of Uniqueness.*—Taking any equivalent pairs from the series of 157, it was found that the operation of adding any two

discrimination of certain individuals seemed to make the notion of a series, which would be linear for all, an entirely hopeless quest. For this reason, except when the inter-personal correlations are high, I prefer scales and schemes of marking based on the assumption of a normal distribution, rather than a rectilinear. Dewar [118] and Eysenck (unpublished thesis) have both taken up the same question in turn, and have come to the same conclusion.

[1] Owing to lack of time, systematic comparisons with all the 157 specimens and all the original 63 observers were not actually carried out. A 'subsample' of the 'sample field' was used instead. This unfortunately makes it impossible to give any clear estimate of the reliability of the final conclusions. Woods and Winter similarly attempted to construct a linear series with children's compositions; and quite recently E. M. John has taken up the same problem with artificially constructed prose passages.

[2] If the reader desires to test them for himself with actual specimens, I suggest that he take the examples of handwriting and drawing given in the L.C.C. *Report* already cited ([41], pp. 371–98), and treat the numerals in the text as the age-labels. (The first postulate cannot be tested, unless the reader can provide himself with equivalent specimens, since only one is printed for each age.)

intervals is, for the majority of the observers (84 per cent.), unique, e.g. the result of adding (8 — 7) or (8′ — 7′) to (7 — 6) yields an interval that appears the same, i.e. (8 — 6) = (8′ — 6).

(ii) *Law of Equality* (*Addition of Equals yields Equals*).—Assuming that all just perceptible intervals are equal, we have (11·0 — 10·9) = (10·9 — 10·8) and (10·9 — 10·8) = (10·8 — 10·7). Then on adding the two sides of the two equations we should have (11·0 — 10·8) = (10·9 — 10·7). As a corollary (11·0 — 10·0) = (10·0 — 9·0), (11·0 — 9·0) = (8·0 — 6·0), etc. On an average, 76 per cent. of the observers agreed with these statements; but with more than two-year intervals, the percentage of those agreeing diminished appreciably.

(iii) *Monotonic Postulate* (*Law of Increase*).—Starting with any pair of equal intervals, e.g. (10 — 9) = (7 — 6) and adding (say) (11 — 10) > 0 to the first, the majority of observers (82 per cent.) agreed that (11 — 9) > (7 — 6).

(iv) *Commutative Law.*—Assuming, according to the definition given above, that (8 — 7) + (7 — 6) = (8 — 6), 92 per cent. agreed that (7 — 6) + (8 — 7) = (8 — 6); i.e. the order of addition makes little difference. With some, however, a slight constant error was noticeable.

(v) *Associative Law.*—Nearly all (98 per cent.) agree that (9 — 8) + (8 — 6) = (9 — 7) + (7 — 6).

In all such inquiries, the approximate verification of the simple, formal postulates proves to be the least interesting outcome. It is clear that the postulates are not wholly inapplicable; but the margin of error is very much larger than that obtaining in fields where they have been accepted almost without question. What is far more suggestive, however, are the cases where the postulates do not apply —particularly the introspections of the subjects and their discussion of the investigator's injunctions, when decision on the postulates is difficult. Flagrant inconsistencies[1] not infrequently occur. But instead of forming

[1] The most striking are the following. (i) It is easy to find a series of scripts in which the consecutive members are indistinguishable according to our statistical standard, while the end-members are perceptibly different. This is like the old paradox relating to 'infinitesimals.' An 'infinitesimal' magnitude was equated to zero; yet a finite sum of such 'infinitesimals' was held to yield a finite magnitude. With 'fundamental measurements' the most familiar illustration is the schoolboy's game of watching the minute-hand of a clock, which seems to be still in the same position after an

grounds for dismissing the whole proposal, they shed a most instructive light on the nature of the process by which such judgments are reached, and thus help to indicate how that process may be made more trustworthy.

All measurement is in some degree inexact and unreliable. When we say that a child is x years old or x feet high, that merely means that his 'true' measurement, X, is such that $(x - \frac{1}{2})u < X < (x + \frac{1}{2})u$. The physicist himself is always ready to distinguish between 'orders of magnitude', and to 'neglect small quantities.' Thus the more important task that confronts the applied psychologist is, not so much to prove that his variables obey the postulates of measurement, but to show that the errors entailed by his tentative methods are small [1] compared with the measurements so obtained.

The same principles and procedure may be applied when our interest lies in comparing, not the mental productions of different persons, but the mental characteristics of the same person. In motivating human behaviour, particularly in the social and economic world, an important part is played by what may be called, in a broad sense of the term, *relative preferences*. The recent effort to extend factorial measurement into this new field has met with strenuous criticism. To a large extent the objections urged are much the same as those raised sixty years ago against the 'dismal science' of economics. The factorist is warned that, since the

interval of 10 seconds, and yet after 6 such imperceptible movements, is seen to be at an obviously different place. (ii) Again, given two scripts, 10 and 6, the observer can find a third bisecting their distance, i.e. such that $(10 - 8) = (8 - 6)$, and two more such that $(10 - 9) = (9 - 8)$ and $(8 - 7) = (7 - 6)$. He is now asked to find another script, 8′, such that $(9 - 8') = (8' - 7)$. Then 8′ and 8 should not be perceptibly different. Not infrequently they are. Such inconsistencies merely illustrate the effect of allowing small deviations, lying within the margin of experimental error, to accumulate.

[1] If the research student asks how small, I suggest, to begin with, the rough convention often adopted by the physicist: "two quantities are of the same order of magnitude when their ratio does not exceed 10" (Jeffreys, *loc. cit.*, p. 217). The weight of the copy of the standard kilogram at the Standards Office at Westminster is certified to be 1·000 000 070 ± ·000 000 002 times the weight of the original in the Standards Bureau at Sèvres. But the educational psychologist seems happy when two copies of his unit differ by no more than 10 per cent.

failure of the utilitarians, any attempt to treat pleasures or values as absolute magnitudes must be regarded as "*prima facie* fallacious." Nevertheless, he may, I think, legitimately reply that an absolute or extensive measurement of values is not what he wants: *relative* preferences can still be treated quantitatively, if they are regarded as *distensive* magnitudes. If my guide book tells me that Sir Richard Wallace paid £600 for a painting by Boucher and £550 for one by Greuze, but declined a painting by Manet offered at £300, I cannot deduce that his absolute love for Boucher was more than twice as great as his absolute liking for Manet; but I can justly infer that the gulf between his preferences for the impressionist style and for the style of Louis Quinze was much greater than the minor variations among his fancies for the 'school of pink and pale blue.'

However, purchase price is at best but an indirect measure of what the economist calls 'intrinsic' or subjective value. Hence, it seems desirable to seek a more direct method of measuring the latter. Without putting our examinees to the equivocal test of an actual auction, we can hand them reproductions in postcard form, and so elicit rankings or gradings, which will measure, if the experiment has been properly planned, not indeed their feelings on any absolute scale, but the *differences* between their preferences. Similarly, we can get children to rank their preferences for different school subjects; and so compare their preferences for each with their achievements in each.[1] Formally, it would seem, all such gradings are quite as valid as the more familiar graded judgments on ability or skill.[2]

[1] Burt [29], [69], p. 278. Here again, however, the proper procedure has been the subject of some controversy: cf. Stephenson, *Brit. J. Educ. Psych.*, V, pp. 43 f.

[2] As a means of studying æsthetic preferences, the 'method of choice' is as old as Fechner. Witmer seems to have been the first to apply Galton's notion of ranking (or, as he terms it, 'method of regular arrangement') to the æsthetic field (*Phil. Stud.*, IX, 1894, pp. 96 f.). Cattell's work, however, gave the ranking method its great popularity. What recent critics appear to doubt is, not so much the validity of the ranking method in itself, as the validity of factorizing correlations obtained from such data.

The view of value as a distensive magnitude seems fully in keeping with modern economic theory. "The Utilitarians thought of absolute value as a quantity—a simple sum of values viewed as atoms of pleasure, which they treated arithmetically. Once the atomistic view is departed from, and economic value becomes the expression of preferential *relations*, the measurement of absolute value is no doubt impossible. . . . But the curious thing is that, though absolute value is not conceivable as a quantity, or is barely conceivable as a quantity, economic values are conceivable as quantities, and

So far, we have confined our attention to the measurement of a single, empirical variable only. Granted that we can find a scale for comparing differences in the same characteristic, what scale can be devised for comparing differences in different characteristics? That, after all, is the essential prerequisite of correlation. Evidently, if we are to combine measurements of two or more variables, other assumptions must be added to our list—e.g. that they can all be reduced to some common kind of standard measure, or that their variances can be stated in commensurable terms. If, further, we propose to investigate objective differences in variance, then we must contrive some method of equating the units employed in measuring the variables to be compared. Such problems have proved somewhat elusive in actual practice, but they involve no new principles in theory.

In my view the most appropriate unit is not the standard deviation, but the just-perceptible difference.[1] In dealing, for example, with English composition, after defining the group of observers and the procedure to be followed, it is by no means difficult, though a little laborious, to select a scale of specimens proceeding by steps that are 'just perceptible'; and further experiment will usually show that, in such a scale, any two 'equal-appearing intervals' contain approximately the same number of 'just-perceptible'

moreover are measurable" (S. J. Chapman, *Elements of Political Economy*, pp. 60–1: cf. the mathematical treatment of 'utility' from Jevons and Edgeworth to Bernadelli and Frisch). I may add that, if we take Meinong's *dictum* literally, even 'absolute value' may be regarded as quantity or magnitude. "That is or has magnitude which allows the interpolation of terms between itself and its contradictory opposite" (*loc. cit.*, p. 8). Now, pleasure and its contradictory unpleasure, 'positive value' and its opposite 'negative value,' form a bipolar class; and in principle, provided we ignore the qualitative differences between 'poetry and pushpin,' the bipolar class can be converted into an ordered class in which a series of terms can be interpolated between the two antithetical poles.

[1] I believe that the success of standard measure in actual practice arises largely from the fact that it so often converts crude measures into multiples of the just-perceptible difference or into multiples of the causal difference. As the Weber-Fechner law has taught us, where human judgment is concerned, what we take to be equal additions are additions proportional to the amounts to which they are added. In judging physical types, a tenth of an inch added to the length of the nose makes as great an impression as a couple of inches added to stature. Now, dividing by the standard deviation, like

steps. That also holds good, as we saw a moment ago, for a carefully selected scale of handwriting. Accordingly, on this double basis it becomes possible to compare individual variability in the two subjects. Even with the most generous allowance for error, there can be little question about the result: except among very young beginners, the range of individual variation is far wider in English composition than in writing. This tallies with the general view of teachers [1]: and, as we shall see later on, is fully corroborated, when we examine data obtained by other methods of estimating facility in the two subjects, e.g. if we gauge by speed rather than by quality (cf. [41], pp. 407, 409, 410). In principle, therefore, the task of finding a universal unit for all so-called 'abilities' should be no more impracticable than planning a miscellaneous store in which every article shall be priced at sixpence.

Even so, however, we have only considered the validity of grading or measuring the initial or empirical variables, such as are directly observed. Many writers would willingly grant that actual performances in a single test, or a set of tests, can be measured (at any rate in certain cases and with varying exactitude) on a commensurable, additive scale; but they seriously doubt whether the same assumptions are equally applicable in the case of the hypothetical factors indirectly deduced from the observed performances. "On your own showing," they argue, "a factor is a principle of

dividing by the mean, reduces the absolute variations to approximately the same subjective scale. It is partly for this reason that, in correlating physical measurements by persons to determine types, I have argued that we should first reduce the crude measurements to multiples of the standard deviation for each trait (for criticisms of this proposal see [96], p. 198 f., and my reply, pp. 173 f. below). The use of the I.Q. effects a similar reduction (since the standard deviation of the crude test measurements is again approximately proportional to the mean): but now the unit expresses, not equally *perceptible* differences, but equal *causal* differences (i.e. differences in innate constitution, or, if we prefer, differences in *rate* of growth rather than differences in *extent* of growth).

[1] On being informed that certain psychologists held that individual variability was virtually the same in all subjects, a schoolmistress of my acquaintance at once produced the afternoon's exercises of her pupils. With few exceptions, the handwritings seemed at first sight almost indistinguishable; but the wide differences in literary merit were patent to the most unpractised eye. Similarly, in many other mental characteristics, differences in the range of individual variation seem obvious, once we consider the point, even with unaided observation.

classification; and your classifications are based, not on one type of performance, but on many, not on a single observation, but on a system. How then is it possible to add and subtract a set of complex systems?" In replying, I must anticipate the sequel. Briefly, I shall maintain that any such system can be adequately represented by a *matrix* of figures; and it is then a simple matter to show that the matrices employed fit the necessary postulates. From this standpoint, the very object of factor-analysis is to deduce from an empirical set of test measurements a single figure for each single individual which will plant him on a linear scale for one of a number of independent classifying principles, although each principle of classification embodies a highly complex system.

The remaining objections urged against factorial measurement turn for the most part on the dubious assumption—commonly accepted by the psychometrist himself—that the factor measurement he deduces is a measurement of some isolated entity—an 'ability,' an 'instinct,' a 'sensation,' a 'capacity for sensory discrimination,' or the like. With perfect justice it is urged that no one has succeeded in demonstrating that these ulterior psychological entities conform to the postulates of addition: "we cannot pile up the intelligences of 50 imbeciles to make the intelligence of a single Shakespeare." But to show that literal summation is impossible no more refutes psychometry than it refutes thermometry. Such contentions are like assuming that the thermometer is designed to measure an entity called temperature, and then arguing that, since two temperatures cannot be superposed to make a third, the measurement of temperature is a sheer delusion.

No doubt, the psychologist's conclusions would be of greater value to the teacher and the psychotherapist if he could phrase his results in terms of constant properties—permanent abilities, permanent predispositions, permanent subjective attitudes, and so on. Ultimately, indeed, our theories may lead us to envisage certain derived concepts, stable and enduring, obedient to some law of conservation. True, we cannot pretend to guess what operations such concepts can legitimately be supposed to admit. But

ignorance in this respect alone would not deprive them of their usefulness: for we might be able to establish that, though the operations themselves are unknown, they nevertheless form a 'group'; and thence an analysis on the lines of group theory would still be attainable. Three kindred points may be noted which seem to be overlooked by most critics. The critics insist, as we have seen, that "no one can safely assign a figure to a mental quality or a mental product." In reply it may be said: (1) the requisite process of addition may be defined without introducing the notion of 'figures'; (2) most of the demonstrations on which factor-analysis is based hold good if the law of combination is not addition at all; (3) many of them hold good even when the elements of the matrix are not figures.

However, so long as our efforts at mental measurement remain at an inchoate and tentative stage, it will be wiser to avoid referring the resulting estimates to remote and hypothetical entities, whose mode of interaction is beyond conjecture. Let us keep to the empirical figures and their weighted combinations. Book-keepers and accountants do not wait till the theorists have formally demonstrated that exchange values obey the postulates of addition, and that costs and prices are valid modes of measurement; nor, when the householder receives an invoice from his grocer, does he argue that the pleasures of cheese and chocolate are commensurable neither with each other nor yet with a magnum of champagne, and that consequently the addition exhibited on his bill contains 'a fundamental flaw.'

I am therefore far from supposing that factor-measurements by themselves can yield an adequate picture of the 'living personality' that stands behind the measurable performances. For such a reconstruction a factor-measurement is about as helpful as the barring and metronome-figures are in indicating the changing rhythm of a symphony. A complete set of factor-measurements for a sample population, with no comments and no case-histories, would be as informative as the bare notes on an orchestral score with no marks of expression and no concert-goer's notes. To translate a musical experience into black dots on a number of parallel staves is indeed to 'disrupt a living whole into little

bits in the interest of a quantitative technique': yet that is no reason for consigning all scores to the fire. Provided factor-analysis tells the truth and nothing but the truth, we need not condemn it for failing to tell the whole truth.

Finally, let me insist that measurement is a means and not an end. Measurement is not the goal towards which classification has been groping; nor is classification an out-of-date substitute for quantitative measurement. In complex biological subjects, such figures, whatever be their function elsewhere, are merely a device for making our efforts at systematic classification more rigorous and more precise.

CHAPTER V

THE DERIVATION OF THE CHIEF FACTOR THEORIES

We are now in a position to examine the origin and justification of the various ' factor theories ' that have from time to time been put forward. These more specialized theories have usually been formulated on the basis of correlations between traits or tests rather than of correlations between persons. In this part of the discussion, therefore, I shall keep primarily to researches of the more usual sort : but most of what I have to say would apply to either kind of factor.

(*a*) *The Four-factor Theory*.—What I have called the four-factor theorem I take as fundamental. From this, as may easily be shown, all the other factor theories can be directly derived by omitting factors of one or more kinds and stressing factors of the remaining kinds.

(*b*) *The Three-factor Theory*.—In the majority of investigations every trait is measured once only : to apply each test twice would double the task, not merely of the investigator, but also of the victims who submit to his tests. But if our tests are not repeated, it becomes impossible to distinguish the two kinds of ' unique ' or ' individual ' factors, viz. what I have called the ' accidental ' and the ' singular ' factors (i.e. s_j from e_j : p. 104) ; the two have therefore to be lumped together under one heading, for which the terms ' specific,' or better ' unique,' are usually employed : for $s_j + e_j$ we substitute u_j. This at once reduces the four-factor theory to a three-factor theory.

In reference to mental tests of the kind most commonly employed, the three-factor theorem has been formulated as follows : " Any one concrete intellectual activity may be considered to depend upon intellectual factors of three different orders : first, the *general factor*, common to all intellectual activities, and known usually as general intel-

ligence ; secondly, one or more *special or 'group' factors*, shared only by a limited number of intellectual processes ; and thirdly, *specific or individual factors*, peculiar to each particular test itself" :[1] the theorem may be succinctly expressed as an equation by writing—

$$m_{ji} = f_{jg}g_i + \Sigma f_{ps}p_{ji} + f_{ju}u_{ji},$$

where the symbols and summation have the same meaning as before. In this early formulation I referred primarily to intellectual traits, since in those days factor-analysis had been almost entirely confined to the results of cognitive tests : but at the same time I indicated that precisely the same theorems could be applied to the factorization of emotional, moral, or temperamental traits.[2]

(*c*) *The Dual-factor Theory.*—Here the important distinction to my mind lies between the general factors and the special or group-factors. From a material as opposed to a merely formal standpoint, the significant and fruitful contrast is the contrast between factors of a comparatively

[1] This formulation is taken from my Historical Introduction to *Report on Psychological Tests of Educable Capacity*, Consultative Committee of the Board of Education, 1924, p. 19. However, I first suggested expanding the 'two-factor theory' into a 'three-factor theory' in a paper read to the Manchester Child Study Society in 1909 on 'The Experimental Study of Intelligence' (cf. [17], pp. 94 *et seq.*). The evidence then cited for the addition of a third type of factor was the work at Oxford mentioned below, confirmed by later work on 'higher mental processes' by means of group tests. On this basis I argued that we must "distinguish between (i) capacities applicable in one direction only, (ii) capacities applicable in several directions, and (iii) capacities applicable in all" ; and quoted Carlyle's "favourite antithesis between 'fundamental greatness' ('the truly great men could be *all* sorts of men') and 'varieties of aptitude' ('Nature does not make all great men, any more than all other men, in the self-same mould')." The main distinction, however, is as old as Aristotle : 'Some persons,' he declares, 'are wise in all respects' (σόφοι ὅλως), 'others wise in parts' (κατὰ μέρος. *Nic. Eth.*, VI, vii, 2). Binet, it may be remembered, similarly contrasts 'general intelligence' with 'partial aptitudes,' and just before his death was planning to supplement his scale of intelligence tests with measurements of other abilities.

[2] Cf. *Annual Report Brit. Ass.*, 'General and Specific Factors Underlying the Primary Emotions,' 1915, pp. 694–6, and, for fuller arguments in favour of this extension, see *Character and Personality*, VII, pp. 238 *et seq.*, 'The Factorial Analysis of Emotional Traits.'

wide range, such as general intelligence, general emotionality, and the like, and factors of a more restricted range, such as the verbal factor, the manual factor, the factor which contrasts introverts with extraverts, etc. The former, as it were, state the *genus*, the latter the *species* or the type. As we shall see later on, the initial inclusion of individual or specific factors (as the psychologist uses the term specific[1]) is little more than a prefatory acknowledgment on the part of the theorist that, however much he proceeds to generalize, his tests are after all particular tests and his persons particular persons, just as his initial inclusion of a set of error factors is an admission that all measurements and assessments, particularly those of psychological traits, are approximations only. Both are irrelevant to his main problem. No psychologist, so far as I know, has ever calculated a person's factor-measurement for any specific factor. If, therefore, we hold that the specific factors, like the chance factors, are devoid of all psychological interest on their own account, and may be dismissed as unwelcome intruders whose influence has to be reduced to negligible proportions because it cannot be wholly dispelled, then we are left with two noteworthy kinds of factors only—the general and the group.

(*d*) *The Two-factor Theory.*—In the past, however, it has been not the specific factors, but the group-factors, whose importance—and even existence—has most commonly been denied; and it is the latter that have furnished the chief

[1] In my own earlier writings (e.g. *Child Study*, *loc. cit. sup.*, p. 98 f.), I used the term 'specific' (and later 'special') to designate factors or attributes characteristic of a 'species'—i.e. what are now most commonly called 'group-factors' or 'type-factors.' Owing, however, to the popular use of 'specific' as a synonym for 'peculiar,' the term has come to be applied by psychologists to designate what might have been called, with less ambiguity, a 'peculiar,' 'individual,' 'unique,' or 'singular' factor: but these words, though occasionally employed by certain writers, have been avoided by most, presumably because in colloquial speech they convey the notion of something exceptional or bizarre. Accordingly, I shall here fall in with the present custom, and restrict the word 'specific' to the narrower sense. Thurstone's definition is clear: "By a *specific factor* or ability is meant any factor or ability which is called for by only one of the *n* tests" ([84], p. 54). Thomson, Holzinger, and most other writers now adopt the same meaning; and any change of usage would only create confusion.

topic of controversy. In the earliest attempts[1] to verify the existence of a general factor, the presence of additional group-factors certainly made itself felt; but from their very nature, their influence was bound to be relatively small, and, in factorial work, was almost always overshadowed by the more dominant general factor. On the other hand, in practical work with backward and neurotic children, and in the psychological clinic where children are referred for special examination, and exceptional cases are no longer the exception but the rule, there it seemed impossible to account for the recurrent types of specialized disability, and for the recurrent contrasts of temperament and character, unless factors of a more limited kind were also assumed.

Nevertheless, Prof. Spearman and most of the earlier laboratory workers could discover little convincing evidence for such an assumption. After reviewing the chief factorial studies of his pupils, he writes: "cases of specific correlations or group-factors are astonishingly rare; over and over again they have been proved to be absent even in circumstances when they would most confidently have been anticipated by the nowadays prevalent *a priori* job analysis. Of 'special abilities'[2] . . . there are but the scantiest

[1] For example, my first experiments on intelligence tests at Oxford showed "a small but discernible tendency for subordinate groups of allied tests" (e.g. sensory tests, motor tests, and memory tests) "to correlate together" after the 'general factor' had been eliminated (*Brit. J. Psychol.*, III, 1909, p. 164). The study of sex-differences, and of hereditary differences generally, particularly in the field of sensation, pointed in the same direction [22], [23]. Conclusive evidence, however, was hardly to be expected until the introduction of group tests [20] allowed us to test far larger numbers and so reduce the probable errors. Similar group-factors were subsequently demonstrated both in emotional reactions [30] and in educational abilities [35].

[2] The reference here is apparently to such factors as the 'verbal or linguistic factor' and the 'manual or mechanical factor,' which had been (so I considered) demonstrated in my earlier work on educational abilities and vocational guidance: in a recent *Report* (quoted above) I had just argued that, in addition to the 'two factors,' we seemed compelled to recognize a third, which I there called 'special abilities or group-factors" (*loc. cit. sup.*, p. 19). In disproof of the "assumed special ability for verbal operations," Spearman quotes Davey's work as 'decisive.' On the other hand, he seems to accept the "special arithmetical factor" which had been confirmed by both Rogers and Collar. The 'mechanical factor' he believes to be explicable by artificial training. He has, however, always urged that the phrase 'special

indications. The modern version of the doctrine of faculties has shown itself none the happier for discarding the old name while retaining the old fallacy " ([56], p. 241).

If with Spearman we drop the group-factors, but retain the specific, our four-factor theory is still reduced to two. The outcome is the famous ' theory of two factors ' [24], [28]. We may express it algebraically by writing—

$$m_{ji} = f_{jg}g_i + f_{ju}u_{ji}.$$

Put into concrete terms the two-factor theory maintains that " all branches of intellectual activity have in common one fundamental function " (the general factor), " whereas the remaining or specific elements seem in every case to be different from that of the others." [1]

(*c*) *The Single-factor Theory.*—Suppose, however, we admit that the status of the specific factors is at least as dubious as that of group-factors, then we evidently shall be left with a single factor only. Both before and since the advent of factor analysis, many writers have attempted to maintain that all mental life, or at least all cognitive activity, can in fact be interpreted by a single unifying principle. Several commentators, for example, have contended that,

abilities ' is inappropriate (cf. [56], p. 222). With that I agree ; but, when I first used it (in my *Child Study* paper of 1909), I was endeavouring to find a terminology that would be intelligible to teachers to whom the whole notion of mental factors was entirely new : and the same reason led me to retain it in the Board of Education *Report.*

[1] ' Theory of Two Factors,' *Psychological Review*, 1914, XXI, pp. 101–15. Spearman, it should be noted, does not deny the *possibility* that group-factors may exist ; he merely considers that there is little or no convincing evidence for the empirical *fact.* When two (or more) tests are very similar, he admits that ' overlapping specific factors ' can be recognized. We are tempted to ask why he is unwilling to treat the overlapping part as distinct from the specific parts, i.e. as constituting a separate group-factor by itself ; and the reason appears to be that in his view the so-called group-factors " indicate no particular characters in any of the abilities themselves " ([56], p. 82). Here the word ' overlapping ' describes the fact that a specific factor may overlap more than one ability or test, viz. all those that are similar (p. 223). The overlapping of one group-factor by another group-factor appears to be definitely excluded. On the other hand, the group-factors revealed in my educational tests showed a definite overlapping, usually in ' cyclic ' fashion ([35], p. 59).

I am tempted to suggest that Spearman's more extended form of his

since Spearman's three noegenetic principles must enter equally into all intelligent activities, they are therefore reducible to one—a principle or process which they variously name 'attentive awareness' or 'simple apprehension.' His specific factors, which he had tentatively identified with the localized functions of the cortical areas or cells, were declared to be "hardly compatible with the newer doctrine of the mass action of the brain." [1] And so an uncompromising 'monarchical' hypothesis (to adopt Spearman's own language) was, according to these writers, the true inference to be drawn from his demonstration of the 'Universal Unity of Intellective Function.'

However, I doubt whether many psychologists of the present day would argue that mental activity could be reduced to a single type of process, with no others to supple-

theory, as given in his latest pronouncement (*Brit. J. Psychol.*, XXIX, 1938, pp. 184 *et seq.*), could be reconciled with my own by regarding it as a series of progressive dichotomies. Thus, it might be said, (i) his first and main distinction is between (*a*) *the* one general factor (*g*) and (*b*) all others, i.e. the non-general; (ii) the latter or non-general may then be subdivided into what he calls (*a*) 'correlated' or 'overlapping' specifics (which he would substitute for the 'group-factors') and (*b*) the uncorrelated, non-overlapping, or unique factors (p. 185, lines 2–5); (iii) finally, we can introduce his own "subdivision of the specific factor" (i.e. of the *non-overlapping* specifics) into (*a*) "correct value" and (*b*) "mere random error" (p. 185, lines 27–8).

However, he himself rather deprecates such further subdivisions, saying that we can, if we wish, "divide up an ability into as many factors as we please, all equally true." For much the same reason apparently he considers that "the current measurements of specific abilities—upon which have come to hang the weal or woe of countless individuals in industry and otherwise—are little more than the blind leading the blind" ([56], p. xviii). Those of us who have been engaged in the practical work of educational or vocational guidance would feel, I think, that the criticism is hardly justified by the concrete results achieved. But this is not the place to defend our procedure. I will only point out that the 'specific abilities' we measure to give guidance in industry or education are '*group* factors,' not '*specific* factors' in the narrower sense: they are such things as the 'verbal factor,' the 'arithmetic factor,' the 'spatial' or mechanical factor—in short, 'overlapping' factors such as many of Spearman's followers, and, it would seem even Spearman himself, are now inclined to accept.

[1] The reference is to Lashley, *Brain-Mechanisms and Intelligence* (1929). Spearman, of course, could now fairly reply that, since mass action alone is no longer held to be a sufficient principle by itself, the true conclusion is that the central function alone can hardly suffice.

ment it. As Guilford has pointed out, " attempts to define this supposed unitary ability have signally failed to satisfy," and " all *unifactor theories* fail to meet the test of accounting for the known factors." [1] Yet, although it would now be almost unanimously admitted that a single-factor hypothesis could not possibly cover *all* the correlation tables which the psychologist obtains, such a hypothesis will nevertheless certainly cover some. And conversely it is nearly always possible, by limiting the choice of tests or traits or else by pooling them, to obtain *a* correlation table, which will fit such a hypothesis, and that in numerous fields of psychological inquiry. All that is necessary is to select the sample of tests or traits according to an appropriate plan, viz. in such a way that they shall *all* represent the same single factor, and no *group* of two or more shall represent the same second factor.

Thus, if all the tests selected involve the same type of material and the same type of task, and differ only in the complexity of the problem and the degree of mental organization that each requires, then group and specific factors will have been virtually excluded. Or again, if they all involve the same general type of process, but differ each from each in the special type of material used or in the incidental types of process each requires, then, although there will be a specific factor peculiar to each test, it will not influence the correlations.[2] In either case, we shall have achieved a situation which is so desirable in all experimental science. We shall have *isolated* a varying factor, and shall thus be able to study its variations in independence of all

[1] Guilford, *Psychometric Methods*, 1936, p. 459 (his italics). Cf. also J. O. Irwin, *Brit. J. Psych.*, XXIII, pp. 371–7, 'A Critical Discussion of the Single-factor Theory' (a most instructive paper, mainly concerned, however, with the mathematical aspect of Spearman's views).

[2] These conditions amount to securing that the sample of tests shall be, not random, but representative. The problem of representative sampling is usually discussed in reference to populations rather than to characteristics. But, as may easily be seen on considering the principles underlying the analysis of variance, the two forms of the problem have many points in common. See [109] and J. Neyman, 'On Two Different Aspects of the Representative Method,' *J. Roy. Soc.*, XCVII, p. 558 f. Where group-factors are included, the method of selection is essentially that of 'stratified sampling.'

other influences. No doubt, the isolation will never be perfect, and the independence will always be approximate and never complete: but limitations of this sort pervade even the simplest investigations of physics and mechanics. Accordingly, let us study the consequences of this artificial isolation.

The Single-factor Theorem.—If any table of selected measurements were due exclusively to the operation of one factor, our initial equation would be reduced to an equation with one term only on the right-hand side, namely, $m_{ji} = f_{jg}g_i$. That means that, with n tests, we should have

$$m_{1i} : m_{2i} : \ldots : m_{ni} = f_{1g} : f_{2g} : \ldots : f_{ng} \quad (i = 1, 2, \ldots N)$$

for each of the N individuals. Thus, corresponding pairs of measurements would always stand in exactly the same proportion, and, to borrow a convenient term from matrix algebra, the whole table of measurements would form a matrix of 'rank' one.[1] The converse of this is the single-factor theorem for a measurement matrix: "any $n \times N$ table, in which the rows (and therefore also the columns)

[1] This conception seems to me so important that a concrete illustration may be given for the benefit of the non-mathematical reader. Suppose, as many Art Schools used to teach, that all the measurements of the normal human body ought ideally to bear the same characteristic ratio to the total height, whatever the individual's height might actually be (e.g. head-height $\frac{1}{8}$, trunk-length $\frac{3}{10}$, leg-length $\frac{1}{2}$, arm-length $\frac{1}{3}$, shoulder-breadth $\frac{1}{5}$, waist-breadth $\frac{1}{6}$, of the total height); then, on measuring N persons of varying stature, all the figures should still be in constant proportion, whatever the individual's actual height. If we merely knew the height of the N individuals —65, 66, 67, ... inches, say,—we could deduce the probable lengths of their limbs, etc., by multiplying these figures by the column of fractions; and the $n \times N$ table of physical measurements so obtained would have a 'rank' of one.

For examples of matrices of rank one in hypothetical frequency tables filled in on the assumption of homogeneity, cf. [25], p. 66, Table III and [50], p. 91, Table 18.

The simplest definition of rank is the following. If every row (or column) in a matrix can be expressed as a linear combination of r linearly independent rows only (but no less), then the matrix has a rank of r. From this, the reader with an elementary knowledge of determinants will easily derive the more usual formal definition: a matrix is of rank r when at least one of its minor determinants of order r does not vanish, while all its minor determinants of order $(r + 1)$ do vanish. Thus, the criterion for a matrix of rank one is that all its two-rowed minors must be zero.

are proportional to one another, can always be expressed in terms of a single 'factor,' that is, as the product of a single column of 'saturation coefficients' (one for each of the n tests) post-multiplied by a single row of 'factor-measurements' (one for each of the N persons)"; or, in matrix notation, $M = \mathbf{f} \cdot \mathbf{g}$, where $\mathbf{g}$ is the row-vector denoting the hypothetical measurements of the N testees in the 'general factor.' It will usually be convenient to express these hypothetical measurements in unitary standard measure. We may then write the equation $M = \mathbf{f} \cdot \mathbf{p}$, where $\mathbf{p}$ now denotes the normalized row-vector of factor-measurements for the first and only factor.[1]

Actually, of course, no empirical table of measurements will ever have such a rank: the figures furnished by actual tests will never be *exactly* proportional for all the persons tested. The first step, therefore, will be to seek a hypothetical set of measurements of rank one, which will yield the 'best fit' to the data experimentally obtained.

For the sake of argument let us adopt the simple notion suggested in my previous memorandum [93] that marks given to the ith examinee for the jth test may be regarded as the result of counting up the number of correct answers he has given. We can express this number as a percentage (or better as a decimal fraction) of the total number of marks awarded. Thus, the marginal totals, at bottom and at the side of the mark-sheet, may be taken to indicate (i) the general mark for each child in the ability tested by all the tests, and (ii) the proportion of marks contributed by each of the tests. We might now treat the mark-sheet as a kind of $n \times N$ contingency table, and examine the data along the lines used for testing 'homogeneity' or independence in 'manifold classifications.' On multiplying the marginal totals, each with each in the usual way (Yule's equation 1 [25], p. 64), we shall obtain a matrix of rank one, which may serve to indicate the 'expected' marks—i.e. the marks we should expect each examinee to get on the hypothesis that there was a single general factor only.[2] To test this

[1] Here and elsewhere capital letters denote matrices (e.g. M) and (where confusion is likely) heavy block letters (e.g. $\mathbf{f}$, $\mathbf{g}$) denote vectors, i.e. matrices containing a single row or column only.

[2] This is the procedure I have used in what I have termed 'factor-analysis by simple averaging' (see *Notes on Factor-Analysis. II. Physical Measurements*).

hypothesis we could adopt either of the criteria proposed by Yule, e.g. calculate ratios for the proportionality criterion ([25], p. 28, eq. [3]),

$$\frac{m_{ap}}{m_{bp}} = \frac{m_{aq}}{m_{bq}},$$

or, what amounts to the same thing, calculate the differences between the 'cross-products' in all the available 'tetrads,' as he terms them. "In the case of complete 'independence' the association is evidently zero for every tetrad" (i.e. $m_{ap} \cdot m_{bq} - m_{bp} \cdot m_{aq} = 0$: [25], p. 69).[1] Systematically carried out, either procedure is really a test for a matrix of rank one.[2]

The precisian, however, will now inquire whether a 'better fit' could not be obtained by *weighting* the tests before they are averaged. But what weights are we to choose? In correlational work "the term 'best fit' is used as in the method of least squares: a 'best-fit' determination will therefore be one in which the sum of the squares of the deviations is a minimum, i.e. the standard error of estimate is a minimum" ([47], p. 159). If we accept this convention for factorial work as well, we have at once a well-established principle on which to base a more exact procedure; and, as we shall see later on, what may be called the 'least-squares method' of factor-analysis yields at once, by an easy calculation, a suitable set of weights and proves to be widely applicable, not only for the case of a single factor, but also for the case of many ([93], p. 247).

As in other problems where the principle of least squares is applied, these considerations lead us directly to the matrix of product-sums or covariances. Now, if the measurements matrix has, actually or in theory, a rank of only one, then the matrix of covariances must also have a rank of only one. This is obvious: for, since $M = \mathbf{f}.\mathbf{p}$ and $\mathbf{p}.\mathbf{p}' = 1$, $R = MM' = \mathbf{f}.\mathbf{f}'$.

We thus reach what may be called the single-factor theorem for a covariance or correlation matrix. "Any symmetric table, in which the rows (and therefore also the

[1] Yule's methods were first described in his paper 'On the Association of Attributes,' *Phil. Trans. Roy. Soc.*, A, CXCIV, 1900, p. 257 f.

[2] It appears as such in the chapter on the elementary theory of determinants in most books on algebra.

columns) are proportional to one another, can always be expressed in terms of a single 'factor,' that is, as the product of a single column of 'factor loadings' or 'saturation coefficients' post-multiplied by a row of precisely the same coefficients."[1] Such a covariance table of rank one it will be convenient to designate by the brief and familiar name, a 'hierarchy.'[2]

[1] In more accurate and technical language: "Any symmetric $n \times n$ matrix of rank one can be expressed as the product of a one-rowed matrix (or 'vector') pre-multiplied by its transpose, the elements in the one-rowed matrix being the square roots of the diagonal elements of the symmetric matrix"; or, in symbolic form, if r_{ij} denotes the element in the ith row of the jth column in the symmetric matrix R, i and j standing for any row and column, then $r_{ij} = \sqrt{r_{ii}} \times \sqrt{r_{jj}}$. In dealing with an actual correlation table (as distinct from a covariance table) the chief difficulty arises from the absence of values corresponding to r_{ii}, r_{jj}, which have to be indirectly computed (see below, footnote 1, p. 152). The examples given in Appendix I, Tables I and III, will make the theorem quite clear; a simple formal proof will be found in Cullis, *Matrices and Determinoids*, II, 1918, equation A, p. 134. I may add that it seems convenient to keep the term "saturation coefficients" for the elements of a factorial vector obtained from a correlation table, and use the newer term "factor loadings" for the elements from a covariance table.

[2] It will be noted that the above conception of a 'hierarchy' is in some respects narrower, and in others broader, than that adopted by many writers: one or two of the tables given as examples of a perfect hierarchy by Spearman and his co-workers would not be hierarchies by my definition (e.g. *Brit. J. Psych.*, 1916, VIII, p. 175; *Psychology Down the Ages*, II, 1938, p. 274); on the other hand, they, I gather, would not accept my bipolar matrices of rank one as hierarchical, nor yet the covariance tables which are not correlation tables.

The original and more usual definition of the hierarchy simply required that the correlations could be placed in "an order such that each is greater than any to the right of it in the same row, or below it in the same column" [12] With the small tables previously employed, this could be satisfactorily judged by eye. With larger tables, containing as many as 156 coefficients, and with group-factors tending to disturb the order, a more precise procedure seemed essential [17]. Accordingly, assuming that a perfect hierarchy would obey the product-theorem, the test proposed was that all the residuals should be calculated and shown to be attributable to chance—admittedly a cumbersome procedure. Later Prof. Spearman proposed to test the orders in the rows by correlating the rows (or columns). By this criterion, a table, like Carey's, in which the correlations diminish *arithmetically* would be accepted as a hierarchy: whereas with my procedure it would not. Spearman's final criterion, however—that the 'tetrad-differences' shall all be zero—is virtually identical with the criterion for a matrix of rank one, except that he further implies that all the correlations must also be positive.

If, however, we are to take the weighted errors into account, that will be equivalent to assuming n additional independent error-factors, each specific to a single test. When the initial test-measurements are expressed in standard measure, the test-variances will be unity throughout, i.e. $r_{jj} = 1$. The variance due to the error-factor will then be complementary to that due to the general factor, i.e. $r^2_{js_j} = 1 - r^2_{jg}$. If we retain r^2_{jg} for the leading diagonal, however, the correlation table will still have unit rank ; and will conform to the single-factor theorem.

The Hierarchy derived from the Product Theorem.—Perhaps the simplest explanation and definition of the hierarchical principle is to be obtained by regarding it as a corollary of the 'product theorem' (p. 49, above). Assuming only one factor to be operative, then, as we have seen, " if test a is correlated with a central factor g to the extent of (say) 0·9, and test b is correlated with the same factor to the extent of (say) 0·8, tests a and b will be correlated amongst themselves to the extent of the product of those two correlations, namely, $0{\cdot}9 \times 0{\cdot}8 = 0{\cdot}72$; and with a large number of tests this will obviously produce a symmetrically arranged table in which all the correlations diminish in parallel and proportionate fashion." Thus, expressing the result in symbols, since $r_{ab} = r_{ag}\, r_{bg}$, we have

$$r_{aa} : r_{ab} : \ldots : r_{ap} : r_{aq} : \ldots$$
$$= r_{ba} : r_{bb} : \ldots : r_{bp} : r_{bq} : \ldots$$

" Suppose, for instance, performance a correlates twice as highly as performance b with performance p ; by hypothesis, this is because the central factor g plays twice as great a rôle in a as in b (i.e. $r_{ag} = 2r_{bg}$). Then, for precisely the same reason, the correlation of a with a further performance q will be twice as great as that of b with q." [1] Conversely,[2]

[1] *Brit. J. Psych.*, III, 1909, pp. 159–60, and *Measurement of Mental Capacities*, pp. 11–12, where an illustrative table is calculated in full (reproduced in Table I of Appendix I below).

[2] For those who desire to demonstrate the adequacy of a single-factor or a two-factor hypothesis, the converse is of crucial importance ; and its proof became at one time the topic of much controversy. If, however, we agree that no empirical table is likely to be strictly hierarchical provided the probable errors are kept low enough, then the controversy becomes of little

if R is a ' hierarchy ' in the sense here adopted, i.e. a symmetric matrix of rank one, then, as is shown in the textbooks of matrix algebra, R may be analysed into the product of a single column-vector $\mathbf{f}_1$, i.e. a matrix consisting of one column only, post-multiplied by $\mathbf{f}_1'$, i.e. the same matrix transposed to form a single row.

Once again, no empirical correlation table is likely to obey the requirements of a rank-one matrix precisely. But, before we can decide whether the discrepancies are really indicative of further and more specialized ' group-factors,' we have to discover what is the best fit obtainable with a single general factor alone. In certain cases, we have seen, we can regard the coefficient of correlation as stating a proportionate frequency—i.e. as the ratio of two frequencies, and so analogous to a probability. This suggests a possible line of treatment.[1] Applying the product theorem, we can treat each *expected* correlation as the product of two such proportionate frequencies. To find the latter we may follow the principle adopted in testing the homogeneity[2] of other double-entry tables: viz.

moment. The converse was formally proved by Garnett ([37], 1916, p. 365; cf. also [78]). His argument proceeded by applying the methods of analytic geometry, in a way very similar to that later followed by Thurstone. He concluded that, " were Burt's conditions for a hierarchy satisfied, each of the n qualities tested would be expressible in terms of two independent factors, of which one was specific, appearing in that quality alone, while the other was a single general factor common to all the qualities "; but if these narrower conditions are not satisfied, while Spearman's ' correlation between columns ' conditions are satisfied, then " the differences between the test-measures and a real multiple of an $n + 1$th variable, y, independent of them all, can be expressed in the same way." Garnett himself holds that the conditions for a hierarchy are likely to be satisfied only in very special cases (when tests affected by group-factors are omitted or pooled). When neither condition is completely satisfied, he proposes to introduce ' independent variables that will no longer be specific factors '; thus finally writing—

$$q'_s = (q_s - ky)\,(1 + k^2)^{-\frac{1}{2}}$$
$$= l_s \cdot g + \ldots + {}_sm_1 \cdot z_1 + {}_sm_2 \cdot z_2 + \ldots + n_s \cdot x_s$$

where g, z_1, z_2, . . ., and x_s are the general, group, and specific factors respectively.

[1] This attempt at a logical analysis of the situation was criticized as lacking in rigour: but, granted certain quantitative assumptions, the argument can easily be put into a more rigorous mathematical form: (e.g. [93], p. 281).

[2] " A classification is homogeneous . . . when the principle of division is the same for all the sub-classes of any one class " ([25], p. 71): thus, if we regard the column of marks obtained by the ith examinee as forming one

" the ' expected ' values can be calculated from the marginal totals, so that the total of the ' expected ' values agrees with the corresponding marginal total " (Yule, [25], ch. v), and the same criteria can be applied to test divergences. Accordingly, in my first paper, I proposed that we should take for each test " the sum of its coefficients as measuring its general tendency to correlate " ; and I indicated that, for fitting the observed coefficients with an ideal hierarchy, " theoretical values might be obtained by various mathematical formulæ." [1] Later, where the best possible fit was desired (i.e. one conforming with the requirement of ' least squares '), weighted summation was substituted for unweighted. But in either case the final values were calculated by applying the product theorem.

To reduce the analysis to the lowest possible terms, one further step remains to be taken. Just as we have nor-

class, then the requirement is that the several traits or tests shall yield a sub-classification of his performances according to one and the same general principle, e.g. the tests or traits must all be aspects of general intelligence, but otherwise independent (uninfluenced by any further ' overlapping specific '). " Tests of homogeneity are mathematically identical with tests of independence " (Yule, *loc. cit.*). This principle, however, usually causes some surprise to the non-mathematical student. The explanation is to be found in the technical meaning attached to the words ' independence ' and ' association.' As Yule points out, " the student should carefully note that in statistics the word ' association ' has a technical meaning different from the one current in ordinary speech." If all our tests measure performances belonging to the same general class, they appear to be more or less correlated. But the ' association ' contemplated by the criteria is a specialized association existing over and above that due to their ' homogeneity ' as members of the same general class or ' universe ' ([25], p. 28 : for symmetric contingency tables resulting in this way, cf. Tables A and B, p. 74, and Table IV, p. 70).

[1] *Brit. J. Psychol.*, III, 1909, pp. 160, 163. It will be seen that this account assumes the presence of self-correlation, $r_{aa}, r_{bb}, \ldots, r_{pp}, r_{qq}$, fitting the hierarchical matrix. With my own formulæ these values had to be supplied by inter- and extra-polation. This procedure aroused some criticism ; and an alternative formula was kindly suggested by Prof. Spearman. However, I now consider that the simple summation formula still yields the quickest and (if successive approximation is employed) the most accurate procedure, apart from the more elaborate method of weighted summation. Although, on theoretical grounds, the relative merits of the different formulæ have been much disputed, the actual differences between the figures obtained are usually all but negligible ([93], p. 294). With a *perfect* hierarchy all the procedures commonly suggested yield precisely the same results; in particular, the results of weighted summation are identical with those of simple summation.

malized the vector of factor-measurements, $\mathbf{p}$, so we may normalize the vector of factor-correlations or 'saturation coefficients,' $\mathbf{f}_1$, thus expressing them as the product of a 'factor variance' v_1 and a single set of 'direction cosines' $\mathbf{l}_1$.

We may then sum up the single-factor theorem in matrix notation as follows:

$$\begin{aligned} R &= H_1 = \mathbf{f}_1 \times \mathbf{f}'_1, \\ &= v_1 E = v_1 \mathbf{l}_1 \times \mathbf{l}'_1, \end{aligned}$$

where H denotes a 'hierarchy' or matrix of rank one, E what I have called a 'unit hierarchy' [115], and $\mathbf{l}'_1 \mathbf{l}_1 = 1$. When obtained by the methods described below (Appendix I, Tables I-III), the factor variance (v) and the direction cosines ($\mathbf{l}$) will be respectively identical with the latent root and the latent vector of the correlation matrix.[1]

Three points may be noted in this theorem. First, it is not limited to tables containing positive figures only. Neither the intercorrelations, therefore, nor the saturation coefficients derived from them, need be exclusively positive. This is of special importance for several reasons. It is widely but wrongly supposed that the most convincing ground for postulating a single general factor is 'the almost universal positive correlation among tests of mental ability.'[2] Such an explanation is misleading. A saturation coefficient is a correlation coefficient; and as such may in theory take negative as well as positive values: only imaginary values are excluded (the chief point which differentiates factor-analysis from the corresponding procedure in quantum analysis). Hence, provided the hierarchy satisfies the proportionality equation, half the table (or rather two quarters) may consist of negative correlations; and the table itself will still be explicable as the product of a single factor with real coefficients. The removal of this limitation permits us to analyse bipolar correlation tables, such as result from eliminating the first positive factor (as in tables of residuals) or from testing a population that is homogeneous as regards the positive factor, by means of repeated applications of the same single-factor theorem. In consequence, when the observed table itself departs

[1] I shall continue to use the symbol v_i for the latent roots, instead of the symbol more familiar in matrix algebra, λ_i, because, in factor-analysis, the ith 'latent root' represents the contributory variance of the ith factor—in the present case the variance of the one and only factor.

[2] Cf. Guilford, *loc. cit.*, p. 464.

from the hierarchical form, it becomes possible [1] to treat it as a sum of two or more hierarchies, the first or dominant hierarchy being usually positive throughout, and the rest bipolar.

Secondly, the theorem, as thus interpreted, leads to a vast simplification of the data. We start with an $n \times N$ matrix, i.e. a table of n test-measurements for N persons; on correlating, the product-moment formula reduces them to a symmetric matrix of $\frac{1}{2}n\,(n-1)$ covariances or intercorrelations (the n diagonal variances or self-correlations being, in most cases, not independent data but arbitrary figures depending on our units of measurement); then, if the table is not itself a hierarchy, we can still regard it as consisting of a dominant hierarchy (giving the closest fit to the observed figures according to the principle of least squares) overlaid by one or more relatively unimportant residual hierarchies; and, finally, this dominant or best-fit hierarchy can be reduced by the single-factor theorem to (i) a vector or column of n normalized figures only, weighted by (ii) a single figure, representing the factor-variance.

Thirdly, the theorem implies that, if the figures representing the test-variances or self-correlations (i.e. the figures in the leading diagonal of the covariance or correlation table) also obey the product theorem and so fit the general hierarchical pattern, there will be no universal obligation to assume any specific factors at all. This negative assumption was, indeed, the basis of my original formula for calculating saturation coefficients by simple summation.[2] On the other hand, even if the inter- and self-correlations do not obey the product-theorem as they stand, they can nevertheless be expressed as the sum of sets of partial correlations that do obey it.

Can the Specific Factors be Dropped?—It will be observed that my derivation of the hierarchy and my formulation of the single-factor theorems do not include any mention of specific factors. If, indeed, we regard a correlation matrix as merely a special case of a covariance matrix, then no specific factors seem required. I do not claim that specific factors will never be necessary, but merely that they are not always and automatically necessary.

[1] A formal proof will be given later: cf. [102], p. 189, [115], p. 156.

[2] See Appendix I, p. 474. The difficulty mentioned above (to discover values for the variances) may appear even greater in the case of tables which, for one reason or another, do not perfectly fit the hierarchy. As will be seen from the account given in the appendix, the simplest device seemed to be to insert figures that fitted the correlational pattern, and then if necessary check their accuracy at the end (p. 448).

Moreover, if we accept the main principle on which the two-factor theory is founded, namely, the absence of group-factors, then (as I have elsewhere argued) the specific factors by their very mode of calculation seem to be little more than errors in the measurement of the general factor: for, according to the two-factor formulæ, the 'saturation coefficients' for the general factor are first calculated on the implicit assumption that the variance of each test is the square of its saturation; next, a second assumption is made, namely, that the variance of each test is equal to unity, and the balance is ascribed to a 'specific factor.'[1] But what evidence is there for any such balance, except perhaps that we cannot assume our tests to be perfect tests of intelligence alone? Certainly, we are hardly entitled to treat 'specific *factors*' derived in this way as synonymous with 'specific *abilities*.'[2] After all, if a

[1] Spearman's formula is $r_{as_a} = (1 - r^2_{ag})^{\frac{1}{2}}$ ([56]), p. xvii, eq. 22). The difference between myself and other writers in regard to the importance to be attached to specific factors seems largely to arise from a difference in our procedure when selecting, constructing, or averaging tests. In demonstrating a hierarchy, the adherents of the two-factor theory are ready to take a set of any dissimilar tests; but since each type occurs only once, the specific factor that it contains is, quite rightly, assumed not to disturb the hierarchical arrangement of the coefficients; in estimating the factor-measurement for the general factor, they claim that "the specific elements will neutralize one another." My own tests also involve differences of material. But in selecting the words, figures, facts, pictures, etc., I endeavour to secure that the special knowledge required shall be common to all the children to be tested, so that it cannot operate as a differentiating factor. Hence the variance is not appreciably increased owing to the specific nature of the dissimilar materials, as it presumably is when this precaution is not taken. (The difference, no doubt, is only a matter of degree; but it is sufficiently evident in applying two batteries of the tests to children of the ages intended.) Where, however, two similar tests are used, and any irrelevant influence unavoidably arises, then, instead of pooling the tests, so as to turn a group-factor into a specific factor, I seek to eliminate it by partial correlation. Those who do not adopt these experimental or statistical devices are naturally led to emphasize the influence of the specifics.

[2] 'To measure a person's specific ability,' Spearman gives the formula $_{ax} = r_{as_a}\, m_{ax}$, where m_{ax} is x's measurement with test a (*loc. cit.*, p. xviii, eq. 24). But that would mean that we give all the persons tested the same order of merit for the specific ability as we do for the test itself. With a battery of correlated tests therefore the specific factors would be correlated.

'specific factor' could be identified with some 'specific ability,' it could obviously appear in more than one test of that ability,[1] and so would really cease to be a strictly 'specific factor': it thus becomes, potentially at any rate, a 'group-factor'—i.e. a factor common to a group of two or more tests.

This view is confirmed by the low degree of stability that the alleged specific factors show from one investigation to another. The simplest test for stability is the symmetry of the product matrix obtained by multiplying the correlation or covariance matrices deduced from the successive sets of measurements. For this purpose the principal diagonals should contain the complete communalities—i.e. the sums of the squares of the saturation coefficients for all the common factors, but for common factors only. If the

[1] The case is perhaps a little different in dealing with correlations between persons: to the logician, the 'specific factor' for each person then becomes almost an attempt to state his *principium individuationis*. But in that case the 'specific factor' can no longer be regarded as on the same footing as the rest. The relation of an individual to his class and the relation of a subordinate class to a supraordinate (e.g. of species to genus), though treated as identical from Aristotle to Frege and Peano, are now recognized as relations of entirely different types; and to treat the 'specific' or individual factor (which depends on the first relation) as a 'factor' in the same sense as the group and general factors (which depend on the second) is to commit the old logical mistake.

So far as concerns factors for traits, it would be far better if the terms 'general' and 'specific' always referred to generality or specificity in the one particular set of tests or traits under examination at the moment: i.e. the word 'general' should mean 'common to all *these* tests or traits' (not to all conceivable tests or traits as seems at one time to have been maintained); similarly, the word 'specific' should mean 'peculiar to one of the tests as used *in this set*': (in that case, there would be no temptation to use the phrase 'specific *ability*'). Since, however, the looser usage has become so well established, it might be well to adopt other terms to convey the narrower meaning: elsewhere (e.g. [93]) I have proposed the terms of traditional logic—'universal' (i.e. covering the universe of traits under examination at the moment) and 'singular' (i.e. peculiar to the measurement of this single trait in this one set of data). It is, after all, solely with the latter distinction that factor-analysis is concerned in any particular instance. Only when the characteristics thus revealed have proved stable from one battery to another can we pass to the broader usage; and hitherto this transition has seldom been explicitly proved, but rather slurred over by the ambiguous use of identical terms.

saturation coefficients for the specific factors are also included, then the symmetry is appreciably reduced.[1] And generally the results of all subsequent calculations, based on the correlation matrix, conform much better with general expectation, if communalities alone are included and specific factors excluded. Thurstone, for example, having first expressed the 'factorial matrix' as the sum of three 'components,' $F_1 + D_1 + D_2$ ([84], p. 54), in most of his later deductions ignores the matrices D_1 and D_2, which contain the 'n specific factors' and the 'n error factors,' and, as I have done, bases his arguments on the 'reduced correlation matrix' $R = F_1F_1'$ ([84], pp. 66 *et seq.*).

Thus Thurstone, whose theory is essentially a generalization of Spearman's, begins by including specific factors in Spearman's sense of the phrase: for, like Spearman, Stephenson, and most other members of the English school, he assumes that all tests or traits have precisely the same variance (p. 62). But in practice he confines himself to the extraction of general or common factors only, and would in principle make the number of those factors a minimum (p. 73). Hence, for Thurstone as for Spearman, the specific factor of a test must again be simply the balance left over from the arbitrarily assumed variance when the portion due to general or common factors has been deducted. In fact, the amount of variance due to the specific factor in a given test j (our s_j^2) is explicitly set equal to $1 - h_j^2 - c_j^2$, where c_j^2 is defined as the 'error variance,' and h_j^2 as the 'common factor variance'—usually termed 'communality' for short (our $g^2 + \Sigma p_j^2$; p. 68). Thus the specific variance of a test is not an independently determined quantity at all. It may be added that, on the theory held by Spearman and Thurstone, the simplest type of mental process (one therefore which would have the smallest number of common factors and would usually be the one measurable with the least amount of error) has the same total variance as the most complex type of mental process; it must therefore contain a specific factor having the greatest amount of individual variability in the population tested. Now this is not only contrary to what we should expect *a priori* from the additive nature of variance, but also in conflict with direct observation: for, wherever we can measure a trait in absolute terms, we find that the more complex mental processes nearly always show a far wider range of individual variation than the simpler, and usually involve a larger amount

[1] This is shown in computations made by Woods and later by Davies and Eysenck, who tried the symmetry criterion with both forms of the correlation matrix.

of error (though not perhaps a larger proportion of error) than the simpler.[1]

In the later portions of his book, however, Thurstone inclines toward a very different view. "The specific variance of a test," he says, "should be regarded as a challenge"; and again, although "the complete elimination of the specificity of each test will not be essential in the early stages of the scientific study of human abilities," nevertheless he regards it as "an object of psychological inquiry to isolate an increasing number of abilities until the specific covariance of each important test shall be reduced to a minimum."[2] These statements come much closer to the standpoint of Spearman's critics. Those of them, for example, who prefer a 'sampling theory' of abilities are naturally forced to argue that a specific factor in itself can represent no real or concrete ability, whether or not they admit the idea of a 'general ability.' Both Bartlett

[1] This has determined my choice of the estimated values for the leading diagonal of the correlation matrix, i.e. the 'total variance of a test,' as Thurstone terms it: (cf. pp. 285 and 460). With a group-factor method I should for certain purposes be willing to insert the reliability. With general-factor methods (except for special purposes) I should prefer not to equalize the variances (with Thurstone, Hotelling, Spearman, and Stephenson) nor yet to set them equal to the reliability coefficients (with Kelley), but to take them as approximately equal to what I have called the 'complete communality.'

Thomson has pointed out ([132], p. 131) that both Thurstone and Spearman are in effect maximizing the specific factors, and he regards this corollary as an objection to the principle that the number of common factors should be minimized. I, however, regard it rather as an objection to the prior assumption which he, in common with Spearman and Thurstone, appears to take for granted, namely, that the variance of all the tests or traits must be treated as the same throughout. If, however, the algebraic equations, derived by treating the reduced correlation matrix as a matrix of covariances, are still to hold good without the assistance of a set of specific factors, then measurements that were initially in standard measure would need to be restandardized; and that in turn would alter the covariances. Once again, therefore, we should have to enter on a further series of successive approximations. Fortunately, these additional adjustments are rarely required. In actual practice, I imagine, covariance will not in general be employed unless there is independent evidence as to the objective differences in variance.

[2] *Loc. cit.*, p. 63. Guilford similarly, after describing the views of Thurstone, Kelley, and Hull, adds: "According to this conception *there are no specific factors*: the elements measured by every test would consist of group-factors plus observational errors" (*loc. cit.*, p. 468: his italics). I imagine the writers named would consider this statement a little too sweeping, but it clearly brings out the logical tendency of their own work.

and Thomson have expressed this conclusion. "Any specific factor must be merely a contrast between a person's 'general ability' and his performance in any particular test." [1]

(*f*) *The Sampling Theory.*—Could we accept the universal presence of the hierarchy as empirically proved, and could we agree on the unreality of specific factors, we should be reduced to a single-factor theory. But even so, the simple factorial matrix thus obtained—a single column of saturation coefficients defining the single factor—will not be the sole factorial matrix that will reproduce the symmetric matrix of correlations: for we can post-multiply this simple factorial matrix by any orthogonal matrix, and the product will reproduce the correlation matrix just as well. In particular, as Thomson has ingeniously shown, by constructing an orthogonal matrix from the Spearman saturations, arranged in binomial groups, and using it to multiply the matrix of Spearman saturations, we can

[1] M. S. Bartlett, *Brit. J. Psych.*, XXVIII, 1937, p. 102. G. H. Thomson, *loc. cit. sup.*, pp. 132 *et seq.* Nevertheless, however carefully we select the traits or tests, we cannot suppose that *nothing* is left, peculiar to each one. What, then, becomes of this peculiar element? When (as is usually the case) we are interested solely in common factors, we shall reduce the specific element to the smallest amount we can—if possible, below the level of what is statistically significant. In that case we may usually treat it as part of the residual error. If that is impossible, its fate will depend on the kind of analysis adopted. (i) With general-factor analysis, a specific factor will become a factor which contrasts one test with every other: it will have a positive saturation for that one test and negative saturations for all the rest (cf. [93], p. 307). It thus becomes a bipolar *general* factor: indeed, the tendency of general-factor analysis is to show that no factor is *absolutely* specific to a single test or trait. (ii) With group-factor analysis a specific factor becomes a group-factor entering into a group containing one test or trait only. Yet we cannot determine the specific factor-measurements from that one test alone (as, for example, Spearman seems to have assumed [56], eq. 24, p. xviii): for, apart from extraneous evidence (e.g. that of the reliability coefficient), there is nothing to show how much must be added to the variance to represent it. Hence, even though all its saturations except one are zero, the specific factor still denotes a contrast. When some particular specific factor is of interest to the investigator, he should of course enlarge the group of one by including more than one series of test-measurements (e.g. by repeating the test). In any case, it seems quite unjustifiable to magnify its importance by equalizing all the variances, as the procedures of Thurstone, Hotelling, and Spearman alike require.

derive a new factorial matrix and a new set of factor-measurements admitting of a very simple and suggestive interpretation : namely, that the new variates are composed of a large and specifiable number ($f_{j1}^2 N$) of small and equal components, drawn by random sampling from a larger pool of N such components, ' all-or-none ' in nature : (f_{j1} being the saturation coefficient for the jth test with the first and only factor). These components, it is suggested, form the real ' causal background ' of our tests, and " may be identified on the bodily side " with nerve-cells or " neurone-arcs." [1]

In such a case, however, as I have shown elsewhere, it is equally possible to reverse the argument, and to demonstrate mathematically that any factorial matrix, deduced in the first instance directly from the principles of sampling, will itself behave as a general factor. Indeed, a formal proof is scarcely needed : for (to put it crudely) a homogeneous brain, consisting merely of a very large number of similar nerve-cells, identical in nature and in strength, would obviously be a brain governed by a single general factor, with no group-factors and no specific factors. In short, there is no mathematical difference between assuming only a single factor, varying continuously, and assuming an infinite (or indefinitely large) number of unit-factors forming a single homogeneous ' pool.' A bushel of wheat is still a bushel, whether we call it corn or insist that it is composed of innumerable grains.[2]

I have discussed this point elsewhere ; and there is no reason for repeating the arguments here, since Thomson, it is clear, would no longer maintain that the sampling

[1] G. H. Thomson, ' On the Causes of Hierarchical Order among Correlation Coefficients,' *Proc. Roy. Soc.*, 1919, A. XCV, pp. 405 *et seq.* ; cf. [132], pp. 271, 302. But see below, p. 208, footnote.

[2] *Brit. J. Educ. Psych.*, IX, pp. 190 *et seq.*, *Brit. J. Psych.*, XXX, pp. 86 *et seq.* (a formal proof was appended to the original paper) and p. 209 *inf.* Largely, but not entirely, the controversy between Spearman and Thomson seems to have arisen from a difference of logical standpoint : Spearman is interested in describing abilities extensionally, Thomson in describing them intensionally. That seems evident if we think of the factors as logical abstractions instead of picturing them as concrete entities—as neurone-arcs in the one case, as mental energy in the other. " The difference between the sampling theory and the two-factor theory," says Thomson, " is that the

hypothesis in its baldest form is adequate to cover all the facts, if, indeed, he ever meant to do so. The logical implications of the theory, which are unexpectedly suggestive, I shall take up later on.

(*g*) *The Multiple-factor Theory.*—When we pass from abstract mathematical deductions to concrete psychological demonstrations, it is evident that neither Thomson nor Thurstone would admit that the 'universal presence of the hierarchy' had been empirically established. As an explanation of the facts, therefore, they would neither of them accept a reduction of the two-factor theory to a single-factor theory: they would rather drop the general factor, and reinstate the group-factors. Thomson, it is true, is not prepared to deny all possibility of a general intellectual factor; but he regards such a factor as unproven and superfluous. To explain the appearance of group-factors, he modifies the sampling theory so as to admit the existence of 'sub-pools' within the 'total pool' of elementary components ([132], p. 283); and, once group-factors are admitted, then, he urges, a general factor is no longer necessary, for it is possible to 'produce a hierarchical order' (or at any rate a very close approximation to it) 'by random overlap of group-factors, without any general factor whatever' ([39], pp. 175, 189).

Thurstone, in his more recent work ([122], p. vii), declares that "so far we have not found any conclusive evidence for a general common factor in Spearman's sense"; yet he, too, seems to concede it as a bare possibility. Actually, however, his method of analysis virtually precludes any such

latter looks upon *g* as being part of the test, while the former looks upon the test as being part of *g*." May I rephrase this as follows? If we take a set of tests (or rather a set of processes tested by cognitive tests) and consider them in *extension*, we shall say that each of these processes is included in a wider class which is defined as being cognitive: i.e. the special test-processes are included in the same general class, labelled *g*. If we consider the same processes in *intension*, we shall say that the concept of any particular cognitive process is a complex concept, which contains as its generic constituent the notion of being cognitive, i.e. the generic quality of *g* is included in the specialized concept of each tested process. Thus the two theories really make the same statement, for, if the sub-class is part of the class, the class-concept must be part of the sub-class-concept.

factor. His interpretation of the 'law of parsimony in scientific description' requires, not (as mine does) that each factor in turn should account for the greatest possible amount of the variance, but that (*a*) the total number of factors entering into the whole set of traits and (*b*) the number of factors entering into each single trait should be as small as possible ([122), p. 150 *f.*). He therefore seeks a factorial matrix in which every factor shall have at least one zero coefficient for at least one of the tests: i.e. no factor is allowed to enter into all the tests; each can enter into a limited group alone.

If, then, from the four-factor theory we omit the general factor as unproven, and if we regard the specific and the chance factors as devoid of real psychological significance, we are left with nothing but common factors confined each to its own particular group. Such group-factors, however, will now of necessity be numerous, or at any rate plural, since their overlapping has to do the work of the general factor. The number of positive general factors is in any single investigation necessarily only one; the number of specific factors would, by definition, be equal to the number of correlated tests; but if there is no general factor whatever, the number of group or partly common factors must certainly be more than one, but should on Thurstone's principles be certainly less than the total number of the tests (as we have seen, it should be about two-thirds with a small battery, or less by $\sqrt{2n}$ if n is large).[1] Accordingly, this view, which prefers to look for a plurality of group-factors rather than for a single general factor, has come to be known as the 'multiple-factor theory.' To distinguish it from the earlier views of those who, like myself, are prepared to recognize two kinds of 'common factor'—a positive general factor, common to *all* the tests, as well as numerous group-factors, common to *some* of the tests only (or, what amounts to the same thing, two kinds of general factor—positive and bipolar)[2]—it should perhaps be termed

[1] See above, p. 109.

[2] In earlier writings I used the phrase 'multiple-factor hypothesis' to describe this view, because the test results were expressed in terms of a multiplicity of factors by means of a multiple regression equation. But with

a ' multiple group-factor theory ' ; for what are ' multiple ' are the factors themselves, not the *kinds* of factors.

the ever-increasing growth of factor theories, much ambiguity is bound to arise if each is merely labelled according to the number of factors (or kind of factors) it postulates ; and accordingly it would be better to adopt more informative titles. To American workers, who have entered the field of factor-analysis comparatively late, the names often convey quite different associations from those they possess for older writers in this country ; and to younger students the nomenclature must be highly confusing. For this reason, too, it seems wiser to rechristen what I formerly called the ' Multiple-factor Theorem ' (regarded as a supplement to the ' Single-factor Theorem ') the ' Theorem of Added (or Superposed) Hierarchies ' (see p. 164).

A word is needed on the relation between the multiple-factor hypothesis and what was called above the dual-factor theory. With psychological data, though analysis nearly always reveals a multiplicity of factors, it also produces (with certain important exceptions) a well-marked duality of kinds. Whatever mode of calculation is employed, the most striking distinction is in the contrast between the first or dominant factor, with all its saturations positive throughout, and the subsequent secondary or supplementary factors—whether group-, bipolar, or specific—which show saturations that are partly positive, and partly negative or zero. Essentially, the first or positive factor represents an average ; the other factors, deviations about an average.

This distinction, as we shall see, becomes of special importance when we come to discuss correlations between persons. Yet from one point of view it is artificial rather than real. In certain cases, we shall discover, the positive factor may disappear ; and, if we are familiar only with tables of correlations between traits, we shall be tempted to suppose that its disappearance is exceptional. Actually, I believe, its *presence* is exceptional—due to the fact that our collection of variables is exceptional. If there are no negative or zero saturations in the ' universal ' or first ' general ' factor, that simply results from the fact that, in selecting a group of traits or a group of individuals to form our initial ' universe ' or *genus*, we have excluded all instances that do not belong to that *genus*. Supposing that we had taken a random, unselected group : then those that do not belong to the *genus* would have been represented as well as those that do ; and our general factor would have been turned into a group- or a bipolar factor. If, for example, in testing cognitive ability, we include, not only measures of intelligence, but also measures of stupidity, then Spearman's *g* would show negative saturations for the latter. Or again, if in correlating persons we include animals as well as men, then the factor of general humanity would appear to be bipolar : it would define humanity by stating what was not human, as well as what was. In practical work, of course, we are obliged to start with *some* more or less well-defined group. But in theoretical work this limitation sometimes introduces a needless complication ; and often, we shall find, it is more helpful to regard the purely positive correlation matrix, so constantly found with psychological data, as in theory a positive north-west quadrant cut out from a larger bipolar matrix (cf. [93], p. 287).

The Multiple-factor Theorem (Theorem of Superposed Hierarchies).—Nowadays, however, every factorist, with the doubtful exception of the strict two-factor theorists, would acknowledge that we have to reckon with the possibility, and indeed with the probability, of a plurality of common factors entering into most of our correlation tables. Whether the supplementary factors cover all the tests, or only a group of tests, or in rarer cases none at all or each only one, must obviously depend upon the particular set of tests we choose. But, in the broader sense, we are all multiple factorists to-day.

In this country the view is by no means a new one. Already in 1917 I argued that we "have to recognize a multiplicity of common factors," and proposed to meet the more complex problem thus presented by invoking the method of 'multiple correlation.'[1] On this principle, we may assume that any $n \times n$ table of correlations or covariances can be expressed as the sum of not more than n independent single-factor hierarchies. We may, in fact, generalize the reduction given on p. 153 above, and write:

$$\begin{aligned} R &= H_1 + H_2 + \ldots + H_n \\ &= v_1 E_1 + v_2 E_2 + \ldots + v_n E_n, \end{aligned}$$

where H, as before, denotes a hierarchy (i.e. a matrix of rank one), v the factor variances, and E the latent hierarchies. If, as before, we use the 'method of least squares,' the E's will form a set of unit hierarchies, defined by the equations $\mathbf{l}_i \mathbf{l}_i' = E_i$ and $L'L = I$, where I denotes the unit matrix:

$$\begin{bmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & 1 \end{bmatrix}$$

(For a simple illustration, see Appendix II, Table IVA.)

[1] L.C.C. *Report, loc. cit. sup.*, pp. 53, 56. By the ordinary proof of 'partial correlation,' if 1, 2, . . . n denote n independent common factors, suitably standardized, we have $r_{ij \,.\, 12 \ldots n} = (r_{ij} - r_{i1} r_{j1} - r_{i2} r_{j2} - \ldots - r_{in} r_{jn}) \div k$; and, if we assume that the final residual, $r_{ij \,.\, 12 \ldots n}$, is zero or approximately zero, we may write r_{ij} = sum of the products of the paired saturation coefficients for the n common factors: this gives us a formula identical with the so-called 'cosine law' (see above, pp. 88, 91).

On this ' canonical expansion of the correlation matrix ' all factorial methods, in my view at any rate, ultimately rest. In virtue of this theorem every correlation matrix can be expanded as a sum of weighted hierarchies. It thus plays much the same part in the factorial analysis of product-moment functions as Fourier's theorem plays in the harmonic analysis of periodic functions, where, it will be remembered, any such function can be expressed as a sum of weighted sines. Moreover, in psychological work, as we shall presently discover, no matter how large the correlation matrix may be, the factor-variances successively obtained from it, v_1, v_2, v_3, . . ., nearly always form a series that converges very rapidly. This, indeed, is the reason for the common statements that in psychology the correlation matrix can always be accounted for by a single general factor only, or (as more cautious writers put it) that the correlation matrix always has a rank of one or nearly one. Neither suggestion is quite accurate. But, just as with our ordinary number system we can express any number, rational or irrational, as the sum of a converging series (e.g.—

$$3{\cdot}141\ \ldots = 3 + \tfrac{1}{10} + \tfrac{4}{100} + \tfrac{1}{1000} + \ldots),$$

and then use the first two or three terms only as a round approximation, so here : having expressed the correlation matrix as the sum of rapidly diminishing terms, we need keep only the first two or three terms, discarding all that are within the margin of error ; and these first two or three will give a close approximation to the original matrix. Finally, we may note, it is this canonical expansion that reveals the striking parallel between the ' factor-analysis ' of the psychologist and the so-called ' spectral analysis ' of the quantum physicist, which happens to turn on an almost identical equation (see [115], p. 160).

This particular mode of analysing the matrix of correlations leads to an equally simple mode of analysing the initial matrix of measurements, which, after all, is the table that primarily calls for analysis ; for we may now write ([101], p. 75, [102], p. 177)—

$$M = LV^{\frac{1}{2}}P,$$

where M is the initial matrix of measurements, suitably standardized, L the orthogonal matrix of direction cosines, V the diagonal matrix of factor-variances, and P the orthogonal matrix of factor-measurements for persons. The latter can be calculated by the equivalent equation—

$$P = V^{-\frac{1}{2}}L'M.$$

If we decide to ignore the factors having the smallest variance, on the ground that they have no statistical significance, we are left with a set of measurements which gives (for whatever number of factors is retained) the best possible fit to the observed measurements as judged by the principle of least squares. If the initial matrix M has been suitably standardized, P and L can be deduced by correlating (or rather covariating) either tests or persons: since, for the covariances between tests we have $R_t = MM' = LVL'$; and for the covariances between persons, $R_p = M'M = P'VP$ (see Appendix II, Tables II and III, for worked examples). Any of the standard methods for computing latent roots and vectors can be used to calculate the factor-loadings; and, once obtained, these in turn lend themselves to very simple algebraic or arithmetical manipulations for deducing regression equations, factor-measurements, rotated factors, simple structures, and the like.

We have now reviewed the chief rival theories hitherto put forward; and we may, I think, fairly conclude that all of them—the 'multiple-factor theory,' the 'three-factor theory,' the 'two-factor theory,' and the 'single-factor theory'—are merely *special simplifications of the general theorem of four factors*. How many kinds of factors we shall actually discover in any particular case must depend on what traits and what persons have been selected for examination. In general, and so far as psychologically significant factors are concerned, the broader form of the 'multiple-factor theory'—a 'theory of common factors,' it might perhaps be termed—seems undoubtedly to provide the best working hypothesis. Specific factors (in the narrower sense of the word 'specific'), together with the chance factors of error, we may regard in the main as simply the

incidental and inconvenient consequences of our imperfect methods of measurement. This implies that we must recognize a marked difference of status between the general and the group-factors (i.e. the 'common' factors), on the one hand, as compared with the specific and the chance factors (i.e. the 'individual' factors), on the other. For the former we may seek a concrete psychological interpretation; the latter will have none.

Relative Nature of the Distinctions.—All through these discussions, however, there is one point which psychologists seem generally to have overlooked, but which to the logician will appear self-evident. As I have endeavoured to insist in all my writings, "the differences throughout are principally differences of degree: the 'general factor' is simply the 'group-factor' that is of the most widespread occurrence; and the 'specific factors' are simply the 'group-factors' that are most narrowly limited in their operation."[1] Give me a list of tested traits which are said to be governed by a general or universal factor only such as Spearman's *g*: I can always add one or more tests or traits which do not contain that factor, and so reduce the general factor to a group-factor. Name any specific factor, said to be peculiar to one tested trait alone: I can always add one or more slightly different tests or traits, guaranteed to contain that factor (or the constant elements of it), and so convert that specific factor into a group-factor. Similarly, as Thomson has pointed out, in many of the investigations claiming to show that only a single general factor exists, and no group-factor, the investigators have frequently begun by "pooling all similar tests" (as they put it); so that what their critics would claim as group-factors are reduced to the status of specific factors. Thus *the distinctions between general, group-, and specific factors are formal rather than material, relative rather than fixed.*

The modern logician would be the first to remind the psychologist that genus and species are "not absolute terms, but purely correlative." "The same term may be at the same time a genus to the lesser classes it contains, and a species of the next more general

[1] [48], p. 19; cf. [46], p. 230.

class : by itself no term can be styled a genus or a species." [1] In the same fashion we may say : by itself no factor can be styled general, group-, or specific ; such designations have reference solely to the particular set of tests and traits that have been correlated.

This the psychological factorist too often forgets. He speaks of '*the* general factor,' as if it formed the 'essence' (as the logician would say) of some unique *summum genus*. In the earlier days of intellectualistic psychology the factorist's investigations were concerned exclusively with cognitive tests ; and he was apt to assume that these covered the whole of the mind. Hence his general or generic factor, as the context usually reveals, was a factor which entered into all his cognitive tests, but (though he rarely said so) into no non-cognitive traits. More recent writers would insist that the mind possesses conative as well as cognitive aspects. Consequently, to give a complete description of any individual we must include his temperamental as well as his intellectual characteristics. If, then, we start with a set of observed measurements which include, not only cognitive abilities, but emotional and moral tendencies as well, Spearman's 'general factor' (*g*, as he terms it, commonly identified with general intelligence) will appear to be no longer a general factor common to all mental traits, but a group-factor restricted to traits of a comparatively specific kind, namely, cognitive or intellectual.

Accordingly, so far as the psychological interpretation of the factors is concerned, the most convenient theory to adopt will be the most comprehensive : namely, that which simply states that the mental reactions of our examinees can always be described in terms of *a number of factors of a greater or a lesser degree of generality*. How many *kinds* of factors we are to recognize becomes a minor issue. In principle, as will now be clear, we must regard all factors as group-factors, and treat the general factor and the specific factors as merely extreme and limiting cases ; and in that sense we have to deal with one kind of factor only.

[1] J. Welton, *Manual of Logic*, I, p. 81.

CHAPTER VI

THE DIFFERENCES BETWEEN P-, Q-, AND R-TECHNIQUES

The Inverted-factor Theory.—One further theory calls for mention—the 'theory of inverted factors,' as it has been termed. It stands on a different footing, being not so much a theory as a method or technique. Until recently most psychologists have for the most part confined themselves to factors obtained by correlating tests or traits; and nearly all the theories reviewed above were originally elaborated on this basis. Almost from the outset, however, correlations have also been calculated between persons; and from time to time such correlations have been expressly studied and analysed from a factorial standpoint.

With this modification of the usual procedure, the rôles of persons and traits become interchanged or *transposed*.[1] Thus, when correlating traits by the ordinary product-sum formula, the expression $\sum_{t} x_{1t}x_{t2}$ means—multiply the

[1] To talk of *inverting* the factors or the theorems is misleading alike to the mathematician and to the logician; and the use of the term has prompted a good deal of criticism that is really irrelevant to the principle essentially involved. As I have often pointed out, the theorems required for analysing correlations between persons are not 'inversions' of those required for the older procedure; they are formally identical with them, and materially their analogues. Similarly, the matrix of measurements with which we start is not the *inverse* of, but a *transpose* of, that which is correlated in the usual way (the rows are merely rewritten as columns, and if necessary re-standardized). And to describe the resulting factorial matrices as inversions of each other is incorrect, except in certain special cases. With my own method of calculation, the matrix of factor-measurements, obtained by covariating traits, was described as being (under certain conditions) the inverse of the matrix of regression-cosines obtained by covariating persons; but that was merely because the transpose of an orthogonal matrix happens to be identical with its inverse. This statement seems to have led later writers who adopted the same or a similar procedure to suppose that it rested essentially on an 'inversion' (cf. [130], p. 406).

measurements obtained with test 1 and test 2, and then sum for all the N persons tested ; when correlating persons, it means—multiply the measurements obtained from person 1 and person 2, and sum for all the n tests. With the former, for example, we compare tests for two *subjects*, say reading and spelling, and ask how far the orders of the pupils, Tom, Dick, Harry, etc., is the same or similar for both ; with the latter we compare two *pupils*, and ask how far the order of the subjects is the same or similar for both, i.e. whether Tom, say, is best at reading and spelling and weakest of all in arithmetic and algebra, while Dick's best subjects are arithmetic and algebra and his weakest spelling and composition. Similarly with temperamental assessments or ratings (the commoner field for this procedure), what are chosen for multiplication and correlation are " not the assessments for the same character-quality among a group of different individuals, but the assessments for a set of different character-qualities drawn up for the same examinee " ([53], 1926, p. 65).[1]

There are practical as well as theoretical reasons for occasionally adopting this alternative approach. It is, for example, hardly possible for John Doe to say whether his visual imagery is more or less vivid than Richard Roe's ; but he can nearly always say whether it is more or less vivid than his own auditory imagery. Accordingly, when it is required to determine the ' imaginal type ' of a group of individuals, they may each be " questioned according to a prearranged scheme about the vividness with which he can imagine certain experiences, and required to arrange those experiences in a corresponding order " ([23], 1912, p. 251 ;

[1] It should be noted that, in the earlier inquiries into the comparative merits of the two methods, what was chiefly in question was the experimental rather than the statistical procedure, i.e. the reliability of psychological assessments obtained on this basis, not the validity of the mathematical analysis as thus applied. Tested in the usual way, there can be little question about the superior ease and consistency of this mode of judgment, at any rate for all important or well-marked cases : obviously " to pick out the strongest and the weakest points in a given individual should be an easier task than to compare the strength and weakness of a given characteristic in a number of different persons " who are supposed to form a typical sample of the ordinary population—which is what the more usual procedure requires ([53], p. 65).

cf. [116], pp. 347 f.). Each man's deviations from the average order in regard to visual, auditory, or motor imagery will then obviously indicate whether he is a marked visualizer, audile, or motile, or whether he simply belongs to the same neutral type as the average person. If by factor-analysis or other means we have already obtained characteristic orders for the three special types, his correlations with those orders will serve to measure the extent to which he resembles the typical visualizer, audile, or motile. We may call the average order for the general population the 'general factor for persons'; and the secondary orders for particular groups or 'types,' 'group-factors for persons' or more briefly 'type-factors.'

In such an inquiry the more usual method of correlating the data by tests or traits instead of by persons—'horizontally' instead of 'vertically,' as Stern once put it ([21], p. 17)—would yield unreliable results, since there would be no means of equating one man's notion of vividness with another's. And generally, wherever the assessor is the same for all the traits of the same person, but different for different persons, there the method of correlating by persons would seem the more trustworthy procedure.[1] This was, in fact, one of my chief reasons for adopting it in the researches just described.

In an ordinary academic research, the investigator can usually get a single observer (e.g. the class teacher) to assess every person in the group he is studying. But for investigations in a vocational institute or an educational clinic, where almost every case comes from a different school, and has therefore to be reported on by a different informant, it appeared necessary to adopt the alternative approach. The method itself was regarded merely as a supplementary device, with manifest limitations and defects, and not in any way as constituting a 'new technique.' Nevertheless, in our practical work at the Psychologist's Department of the London County Council and in the Vocational Department of the National Institute of Industrial Psychology, and later in more theoretical investigations at the London Day Training College and University College, my fellow-workers and I found it both practicable and useful. In particular it appeared specially suited for numerous incidental

[1] Cf., for example, [23], p. 251, [53], p. 66.

problems in education and vocational guidance—e.g. for studying the preferences of children and adults, for assessing the agreement among school or university examiners, for estimating the reliability of psychological observers, for analysing the nature of alleged temperamental and clinical types, and generally for all those inquiries in which the performances of the persons examined had to be compared with a subjective rather than an objective standard, i.e. was itself a set of personal reactions or judgments. Here, however, we shall be concerned, not so much with the concrete results, as with the exceedingly instructive controversies to which the proposal has lately given rise, and with the light those controversies seem to shed upon the whole nature of factor-analysis.

Criticisms.—The legitimacy of the procedure has recently become the subject of some debate. Several competent authorities have strongly criticized the extension of correlation to problems of this type; others have argued that it might be developed into a new instrument of research and made the basis of an " entirely new branch of psychometry." Objections, I think, were first explicitly raised by Dr. E. C. Rhodes, investigator to the English Committee of the International Institute Examinations Inquiry, when, in 1935, I suggested applying the method to data collected for the Committee, and so studying the reliability of the examiners' marking along lines we had previously adopted in evaluating tests [41] and in reviewing our own College results.[1] About the same time, Professor Thomson, who had himself used correlations between persons for investigating teachers' marks for essays [55], expressed grave doubts about any wider extension of the method ([87], pp. 75–6). He holds that " probably correlations between persons will be in the general case impossible to calculate " ([132], p. 201); nevertheless, he admits that in certain special cases, e.g. where the examinee is required to rank the ' tests ' (pictures, essays, etc.) in order of preference, such correlations may be legitimately calculated.

[1] The essential idea was to test the presence of a general order of merit, influencing all the examiners in varying degrees, by looking for a hierarchical order among the intercorrelations. This principle had been followed, not only for the Binet tests ([41], 1921), but also for junior county scholarship examinations and university examinations (for the Teacher's Diploma, etc.) with which I had been connected.

1. *The Disparity of Units of Measurement.*—His main criticism [1] is based on the fact that the units of measurement for different tests are generally incommensurable. But that objection, as it seems to me, comes from taking the word 'test' too literally. For purposes of expounding the alternative method it is no doubt convenient to draw a sharp contrast between 'persons' and 'tests.' Yet, after all, what we correlate when we 'correlate tests' are measurements for certain persons' traits; and what we correlate when we 'correlate persons' are in theory measurements for the same traits in the same persons. The only difficulty, therefore, is to select traits and to find units which shall be consistent with the particular form of statistical analysis in view. After all, most examiners, I imagine, would be quite as ready to compare the same examiner's marks for different school subjects (or 'tests') as they would be to compare different examiners' marks for the same subject; and every teacher is continually contrasting the different abilities of one and the same individual: 'John is much better at Latin than he is at French'; 'Joan did not do so well in the arithmetic tests as she did in reading and dictation.' If, to take Thomson's own instances of incommensurable measurements (p. 201), the marks obtained in an 'analogies test' were insuperably disparate from those obtained in a 'dotting test' (i.e. if marks for such tests could not possibly be transmuted into commensurable units) how could we ever cross-multiply the figures to obtain the product-moment correlation between those tests?

But, it may be said, in correlating a couple of tests we begin by averaging marks for the same test, whereas in correlating persons we must first average marks from a number of different tests. How can we average such figures if the units of measurement are dissimilar? As a matter of fact, the answer is still the same; and Thomson himself really supplies it in an earlier chapter of his book ([132], pp. 114 f.). There he proposes to find a weighted average for four such heterogeneous tests as picture completion, geometrical analogies, a reading test, and the Stanford-Binet test: his method is—first "we reduce the scores to standard measure" (p. 116). Once the arbitrary measurements furnished by the tests have been changed to standard measure, we can average the different tests with or without an additional weighting. That is all that is required for correlating by persons. This or its equivalent was, indeed, the plan suggested in my *Memorandum* for the Examinations Inquiry Council, namely, "first to reduce the crude scores to terms of the

[1] With his discussion of an important side-issue—the so-called reciprocity principle—I shall deal in Part III.

same unit, e.g. the standard deviation or ranks " ([93], p. 276) ; and one or other of these methods of scaling have formed the basis of most of the assessments correlated in this way.

The real difficulty, as it seems to me, arises, not so much from the fact that the *units* may be incommensurable as from the fact that the *traits* may be incommensurable, i.e. they may be disparate and unrelated items, which we treat as forming a class, when actually they form no class at all. Thus, as an example of factor-analysis applied to correlations between persons, it is suggested that we should take measurements of " height, trunk-diameter, arm-length, leg-length, circumference of neck, breadth of nose, and length of little finger, and so forth " ([96], p. 198). These measurements would all be in inches or centimetres throughout. So far as the unit is concerned, therefore, we could certainly average the length of arm, leg, little finger, etc., for each separate person, and then calculate the standard deviation, and finally correlate the measurements by persons, as the passage quoted then instructs us to do. But would such averages, standard deviations, or correlations have any value or meaning ?

Most people, I fancy, would say no. I would rather say, we cannot answer until we know the problem at issue. The problem for which this procedure was actually proposed is the determination of physical types[1] ; and for such a purpose it seems obviously faulty. But had the problem been to study the resemblance in bodily shape between persons of different size (e.g. between children and adults, or between an ateleiotic dwarf and a normal adult), then, though I should have made a slightly different selection of traits, the procedure itself would seem valid.

I hold, therefore, that the first essential is to show that the traits selected for comparison form an intelligible sample of an intelligible universe or class, and that this class constitutes an ordered class, in the sense defined above (p. 118). Even then, before we could proceed to calculate product-sums, covariances, and correlations, further requirements would have to be fulfilled at each stage of the

[1] The variations in the small features, such as length of nose, breadth of head, which may be quite as important in determining physical types as variations in height, etc., would be swamped by the variations in these larger features. In my own studies of physical and temperamental types, I had correlated traits in the usual way and had also correlated relevant traits by persons after first reducing each trait to standard measure. Stephenson, however, argues that his ' system 2 ' (Q-technique—" reducing the crude measurements to standard measure for each *person* ") is the only proper system for eliciting " individual differences *in type* " ([96], p. 205). My own method (' system 3 ') he rejects for reasons which we shall examine in a moment.

calculation. What the requirements are may be gathered from my discussion on the postulates of measurement in the preceding section.

Difference of Aim.—Where the preliminary process of averaging or ranking by persons is legitimate, it is evident that it must automatically eliminate from further consideration any differences in average or general level that distinguish one person from another. Thus, if we wish to know whether finger-length varies with height, we ignore the fact that fingers are small objects and statures large objects, the difference in the absolute size of the two traits being irrelevant to the comparison. Similarly, if we wish to know whether the proportions of the dwarf correspond with those of the normal adult, we ignore the difference in body-bulk of the two persons. And generally, in correlating traits, we virtually eliminate the general factor that would have been obtained by correlating persons; and, in correlating persons, we virtually eliminate the general factor that would have been obtained by correlating traits. Even if we do not actually 'standardize' the measurements, i.e. if we rely on covariances or product-moments, we are still forced to deal with deviations or with differences, that is, with relative measurements, not with absolute; and that means that with either procedure we still discard a factor that the alternative procedure would preserve, namely, the so-called general factor. In my view, however, *this is the only essential difference between the two sets of factors* derived by the two opposite procedures. The remaining factors (the subordinate, 'bipolar,' 'group-,' or 'type-' factors) will in theory be essentially the same whether we start by correlating persons or by correlating traits.

If this view be correct, the difference between the two modes of approach lies rather in a difference of aim or interest than in a difference of technique. We may correlate persons for one or other of two reasons. First, we may be interested in the general factor for persons, and prefer to investigate this on the basis of a weighted average rather than an unweighted.[1] Secondly, we may be interested, not

[1] An explicit example may make this clearer to those who are not already familiar with the general procedure. In the Binet-Simon tests the scheme of

in the group-factors for tests, but in the group- or bipolar factors which lead us to classify persons into different groups or antithetical types.[1] Nevertheless, as I shall show later on,

measurement assumes that the order of difficulty of the several tests is approximately the same for all. In standardizing by age-groups, therefore, we may begin by averaging the orders obtained by different examiners who have used the scale. But this proceeding suggests a number of prior questions. What if the orders obtained for different examinees (e.g. normals and defectives, boys and girls) or the orders furnished by different examiners vary so widely from each other that a single average order would be meaningless? Can we plausibly show that there is really one ideal order, acting as a general factor, common to every examiner and to every examinee, predominating over, though no doubt disturbed by, other irrelevant influences? And in that case, would it not be better to weight the orders of each examiner according to his agreement with this ideal? To answer such questions it is natural to compare the orders obtained from different examiners and from different children or types of children. On correlating the orders of examiners, for example, the hierarchical character of the table ([41], 1921, Table V) " strongly suggested a single ' central factor,' presumably the ideal order," while the coefficients themselves clearly indicated that the improvement obtainable by the best weighting would be so slight that the empirical order heading the hierarchy could (as so often occurs) be accepted as a reasonable approximation to the ideal. On correlating orders for examinees, it was found that the general order of difficulty accounted for about 70 per cent. of the variation, the remaining portion being attributable to irrelevant factors, such as differences in sex, social status, verbal facility, and the like (*loc. cit.*, pp. 136, 195; cf. [29], 1918, pp. 8, 23; *Eugen. Rev.*, XXX, pp. 255–60).

[1] Once more an example will make the procedure clearer. In my book on *The Young Delinquent* (Appendix I, pp. 614–16) are four typical portraits, chosen partly because the children depicted represent a ' contrast in temperament.' After studying their physiognomy, pose, dress, and letters, let the reader assign to each of them a mark for the strength of McDougall's eleven ' primary emotions ' (enumerated in a previous chapter of that book: the exercise is suggested, and a schedule of traits appended for the purpose, in the monograph on *The Measurement of Mental Capacities*, 1927, pp. 29–30). If he considers each child in isolation, the reader may find it difficult to say whether ' Arthur ', e.g., is above or below the average of the general population for ' Joy ' or for ' Sex ' and by how much, but he will find it comparatively easy to decide whether Assertiveness is more marked in him than Submissiveness, Anger than Fear, and so to group or rank his various traits according to their strength. An earlier page of the book (p. 515) indicates the comparative strength of these same traits in the two commonest temperamental types—in the repressed or inhibited ' introvert,' and in the expansive or aggressive ' extravert ': the former is distinguished by the greater strength of such emotions as fear, submissiveness, tenderness, sorrow, and disgust (roughly in that order); the latter by such emotions as assertiveness,

so long as every person is assessed by the same observer, the best and quickest method is still to proceed by correlating traits, and to calculate factor-measurements for the several persons (so-called 'T-technique')[1]; if, however, each person is assessed by a different observer, and particularly if the persons are fewer than the traits or tests, then the direct method is also the most practical, namely, to proceed by correlating persons (so-called 'P-technique').[1]

2. *The Unnecessary Multiplication of Factors.*—The view I have here put forward has been criticized still more strongly by my colleague, Dr. Stephenson. His chief criticisms have

sociability, anger, curiosity, joy, and sex (roughly in that order, the reverse of the order in which they appear in the typical introvert). We can thus *correlate the ranking for each individual with the ranking for a 'standard personality'* as I called it—i.e. with a hypothetical set of marks representing (in this case) the ideal extravert type: a positive correlation will then indicate a tendency towards extraversion, a negative correlation a tendency towards introversion, and the numerical size of the correlation the closeness of the resemblance.

Copies of a graded schedule of traits have been regularly used for such ratings and type-correlations, not only in class exercises with my students (with the portraits on the lantern-screen), but also to obtain temperamental estimates both of my clinical cases from teachers and of my own students from each other. The 'reliability' of such rankings is unexpectedly high (cf. Burt and Spielman [53], p. 66). For exact research the method may seem rather crude; but for the working purposes of educational and vocational guidance it has proved eminently serviceable (see below, p. 426 f.).

If, instead of merely ranking the traits in order for each individual, the reader endeavours to assign marks, he can correlate the marks by traits as well as by persons. He will probably agree that this is a less reliable proceeding; but, even so, I fancy he will have little difficulty in convincing himself that the correlation of each person with the 'type' is roughly proportional (*a*) to the factor-saturation obtained by factorizing the correlations between persons, and (*b*) to the factor-measurement obtained by factorizing the correlations between traits. However, these are points which I shall take up in Part III.

[1] I plead guilty to using such shorthand phrases as P-axes, P-factors, and even the P-method; but I do not care for the label P-*technique*, because, in my view, there is no essential *technical* difference between the analysis by persons and by tests. P-factors and T-factors are convenient and colloquial laboratory abridgments for 'factors obtained by correlating persons' and 'factors obtained by correlating tests.' By R-technique is to be understood a special form of so-called T-technique, namely, 'that which analyses correlations between tests (or traits) by the application of the Spearman two-factor theorems' (Stephenson).

been brought together in a monograph on *Type-analysis*, to be published shortly; and I have to acknowledge his kindness in communicating them to me in a series of letters and memoranda. Like Dr. Rhodes, he was at first disposed to reject the method outright. The difficulties he originally felt are best summed up in an early note criticizing Dr. Dewar's first account of her investigation and commenting on a projected scheme for students' researches on similar lines in the laboratory to which we were then both attached. His objections are worth stating explicitly, not only because they give the reasons for the important modifications that he was subsequently led to suggest, but also because they voice the doubts so often raised by those who are new to the proposal.

In a number of experiments carried out with the help of Miss Bulley and Miss Pelling, I had endeavoured to show, by correlating the 'marks' given by different persons to sets of pictures, vases, colours, etc., that "there was one general factor influencing the artistic judgments of all," and in addition several less obvious factors, producing more specialized types of appreciation (somewhat similar to Bullough's types, and apparently related to more general temperamental tendencies [75], 1933, p. 292). These conclusions seemed confirmed, not only by minor investigations by earlier students, but also by a long research carried out by Dewar with school children, art experts, and unselected adults.

Stephenson, however, held that the statistical procedure adopted in all these inquiries (a method of multiple factorization applied to the correlations between persons judging) was "at once misleading and futile" and "in conflict with the principles established by Spearman." To start by correlating judges, instead of items tested or judged, was, he argued, "misleading," because it suggested that more factors entered into the tests than we know to be the case. "With a hundred persons to be correlated, the number of intercorrelations will be enormous; and we shall consequently produce a spurious increase in the number of factors": when we correlate by tests "we know from Spearman's studies that there will be only one general factor in each battery," and no more than five or six "over the whole range of the mind." Moreover, the procedure "treats test-measurements, obtained in disparate units—a mere heap of irregular and unrelated items—as a normally distributed statistical population." Secondly, he considered the method to be

" futile," and to " entail a vast amount of unnecessary labour," because the general abilities revealed and " your ' subjective ' and ' objective ' types " can all be reached, " as you have already shown, by the direct and well tried route of correlating tests or traits. As Spearman himself has pointed out, the vital thing in studying types is simply to determine what *trait* goes with what " (cf. [56], p. 54).

Now in many instances, as I readily own, to correlate the large number of persons instead of the small number of traits may entail not only a lengthy but a needless détour. But for the two purposes I have mentioned this course seems all but indispensable. Since I hold that a multiplicity of factors is also reached by correlating traits, the mere multiplicity of factors that results from correlating persons does not shake my faith in the validity of the method. Yet it would be a mistake to imagine (as so many others have also done) that more factors must necessarily emerge just because there are more correlations : the rank of a correlation matrix derived from the one and same set of measurements remains the same whether we correlate by persons or by traits. As for the disparity of units, that can be met, as we have just seen, either by standardizing the measurements for each trait (e.g. when dealing with traits that are normally distributed among the persons correlated) or by standardizing or ranking the measurements for each person (e.g. when dealing with items that have an approximately normal or linear distribution in the sample judged, such as essays by an unselected batch of candidates or a series of specially selected test items spaced out at approximately equal intervals).

3. *Confusion between Variable and Population.*—A third criticism, perhaps the commonest of all, is that the proposed method of analysis " confuses the distinction between variable and population." This has been urged, not only by Stephenson, but by several other critics.[1] It is, however, not an objection that should be brought against the procedure itself, but only against the selection of measurements to which the procedure is applied.

If we take the term ' population ' literally, to denote a *human* population (as, of course, it did in the early days when the statistical terminology was taking shape), the objection may seem plausible. The ' population ' from which the sample figures are drawn is con-

[1] Stephenson, *loc. cit. sup.*, cf. [136], p. 20, [138], p. 275.

ceived as a collection of individuals. The 'variables' are the varying attributes or traits that can be predicated of them; they vary together because they inhere in the same individuals. But how is it possible to reverse the descriptions, and talk of covariability not between traits but between persons? Surely, the student asks, we cannot call the traits a population, and treat the persons as variables predicable of the traits?

The answer is that the concrete context which suggested the current nomenclature is really irrelevant to the abstract argument. What we call a 'person' is simply a set of trait-measurements, considered by columns instead of by rows. As so often happens when new notions are introduced into mathematics at the outset of a fresh line of work, the terms have been generalized. Since nearly all the correlational researches of the psychologist hitherto happen to fit the literal sense of the words, he forgets that the words themselves (as is shown by half the examples in a statistical textbook) are no longer tied down to their original meaning. Both 'population' and 'variable' denote classes. The use of two names indicates a cross-classification. The application of the different names shows which set of values we are, for the moment, taking as predetermined or constant, and which we are treating as indeterminate, i.e. variable or determinable at will.[1]

[1] The illustration which gave rise to this criticism was the proposal in my original memorandum to apply both methods of analysing variance and both ways of calculating correlations to a sample mark-sheet, containing marks awarded by 6 examiners to 15 candidates in a school certificate examination [93], cf. [134]. To correlate the marks both by examiners and by candidates would be, we are told, to confuse 'population' with 'variables': but we are not told *which* should be regarded as the 'population'—the examiners or the candidates; nor are the terms 'population' or 'variable' explicitly defined.

As I myself have employed these terms in previous articles, I ought to justify my own usage. "The idea of a population is not to be applied only to living or even material individuals. If an observation, such as a simple measurement, be repeated indefinitely, the aggregate of the results is a *population* of measurements" (Fisher [50], pp. 2–3). To avoid the misleading associations that attach to the word 'population,' Yule prefers to speak of the 'universe.' "A sample from a universe is a selected number of individuals each of which is a member of the universe. A universe, like any class, may be considered as specified by an enumeration of the attributes common to all its members" ([110], pp. 25, 332; or, I would add, by enumerating the members: cf. Russell, *Principles of Mathematics*, p. 69). "A variable is a symbol which represents any one of a class of elements. The elements of the class may or may not be numbers" (Young, *Fundamental Concepts of Algebra and Geometry*, p. 193; cf. Russell, *loc. cit.*, p. 89).

I have in front of me a mark-sheet giving marks gained by all the boys in a certain school in all the subjects taught at that school. The headmistress tells me she is interested in the way two *subjects* vary together, particularly the way composition seems to depend on reading. How, she asks, can one predict the mark in composition for any child from his mark in reading? To solve her problem I must correlate the two tests, and compute the appropriate regression equation. But now, she tells me, she is also interested in the way the marks for two *pupils* often correspond: Richard, the mentally defective, like Hugh, his normal twin, is much better in manual subjects than in verbal, and worst of all at arithmetic. How (we may suppose her to inquire) can she predict Richard's probable mark in any test from Hugh's? To solve this second problem I must now correlate the two persons. Moreover, I claim (and here I differ from Stephenson) that this can be done quite legitimately with the same set of data. Provided the two types of problem are kept distinct, I see no 'inconsistency' or 'confusion' in changing the standpoint, transposing the labels, and treating the former variables as constants, and the former constants as variables. The incidence of the word 'any' indicates which of the two classifications is taken for the moment as providing the 'variables,' and the transverse classification then represents the 'population' or universe. If I seek to express *any* value of a given test-performance as a function of another given test-performance, then the two *tests* are the 'variables' (dependent and independent respectively) and the aggregate of the values for the different *children* make up the 'population.' If I seek to express *any* value of a given child's test-performances as a function of some other given child's test-performances, then the *persons* are the 'variables' and the aggregate of the values of their performances in the different *tests* must now be called the 'population.'

I do not deny that, if we are to use the same twofold table for the two different problems, then that table must first fulfil certain conditions. As I myself have insisted, several attempts to apply correlation to 'persons' have been invalidated, because these preliminary conditions have been disregarded. The point most easily overlooked by the psychologist is that his set of tests or traits are merely a sample of the total 'population' of traits belonging to the same class, just as his group of persons are merely a sample of the total population of persons belonging to the same class. But an invalid application does not invalidate the procedure itself. The situation is identical with that which occurs in the analysis of variance. Some tables can be analysed for one criterion only;

others for two. When Fisher analyses the variance of rainfall first by hours and then by months ([50], p. 222 f.), exactly the same treatment is accorded first to the rows, i.e. to the hours as 'variables,' when the months form the 'population'; and then to the columns, i.e. to the months as 'variables,' when the hours form the 'population.' Will Dr. Stephenson argue that Fisher has here "confused the distinction between statistical variables and statistical populations"?

However, I need not press my arguments further, because Stephenson himself has largely withdrawn his objections, or rather has sought to show that they need no longer be fatal under the conditions he would prescribe. He himself now talks of traits as forming a 'population' and of persons as being 'variables' [98], and is prepared to correlate persons as well as traits ([96], [97], *et al.*). He still maintains that the two modes of approach are "essentially opposed both in principle and in results"; but he no longer considers this to be a ground for rejecting the correlation of persons, but rather for "breaking with the conceptions of the old R-technique." And in his paper on the 'Foundations of Psychometry' [96] he seeks to show that there need be no inconsistency in such a breach, once we recognize that correlating persons "has nothing to do with the particular branch of research for which R-technique (as developed by Spearman) was and is the proper procedure," and realize that the change of direction takes us "into a new world of factor-analysis, with new problems, new principles, and new premises of its own."

Q-Technique.—What, then, is this 'new department of psychometry,' and what are the new principles that it requires us to accept? The new branch, it is suggested, may be termed "Personalistics." Its aim is "to deal with whole personalities and with 'total situations,'" as distinct from the study of isolated abilities to which psychometry has hitherto been confined. "Though person-factors had previously been used," he explains, "the wider implications of such factors had been almost entirely ignored" ([138], p. 276); and these implications, properly appreciated, lead to a "Gestalt-like view of personality, immune from the criticisms previously brought against correlational analysis by clinical theorists." The special principles he proposes constitute (so he claims) 'an entirely new technique,' which he terms the 'theory of inverted factors' or more

briefly ' Q-technique.' Q-technique is defined as " the factor study of persons as variables." Like P-technique, it thus involves " a reversal of the usual rôles of persons and tests in factorial work." " Prof. Burt," he continues, " has suggested one way of effecting this reversal: but it leads to no new principles or premises in mental-test theory. The function of a door is not changed by hanging it back to front: what I want of the concept of correlating persons is a new door, not a badly hung back-to-front door. . . . If Q-technique is to be of any importance on its own account, it must seek new fields of endeavour in psychometry " ([136], pp. 33–4). " P-technique," in short, " is still R-technique; but Q-technique lies poles apart from either."

There is much that is attractive in these various suggestions; but, apart from the new terms, are they quite so novel as they sound? As Davies and others have pointed out, it is not at first sight easy to see where precisely Stephenson's ' technique ' departs from that of previous workers who had based their factors on correlations between persons rather than tests. Let us note the points of agreement first of all. To begin with, he now accepts the use of standard measure (or some such equivalent) as a means of removing incomparability of units; and in his first experiment on correlations between persons ([92], p. 21), he actually adopted the standard scale and frequencies I myself had drawn up ([35], p. 49, column 6). In his ' Introduction to Inverted Factor Analysis ' [98] he begins by reproducing my matrix formulation of the problem and describing the methods of standardization I had proposed: but then he introduces a novel concept—that of ' significance,' which, as he subsequently claims, ' affords an entirely new basis of quantification.' In his later, formal algebraic proof he adopts my own proposal that, for theoretical work, ' unitary standard measure ' should be adopted instead of ' ordinary standard measure ' ([96], p. 197, eqs. (2) and (3); cf. [93], p. 272): but once again this leads him to draw a sharp and instructive distinction between the alternative ways in which such standardization may be applied. In spite of these new suggestions, however, we seem so far to be in pretty close agreement.

On the other hand, with a view to surmounting the objections to which (as he believed) our earlier work was open, he proposes ' radical modifications ' in both the experimental and the statistical procedure that had previously been employed ; and it is on these modifications that he lays the greatest stress, since ' neglect of them in the past has vitiated nearly all work on correlating persons, and obscured its new possibilities.' The experimental procedure he proposes to alter by substituting " homogeneous tests " for the heterogeneous test material, which was used by Bulley, Dewar, and myself, and which after a preliminary trial he felt bound to discard : e.g. instead of " a sample of 50 picture postcards representing all degrees of artistic merit from reproductions of masters to the crudest birthday card," he substitutes " 50 postcards of Japanese vases all of approximately equal merit." With this altered method of selection, he urges, the tests will no longer constitute ' a heap of irregular and unrelated items,' but may legitimately be treated as a ' normally distributed population.' The statistical procedure he would modify by substituting Spearman's two-factor theorems (in ' inverted ' form) for what he terms my ' multiple-factor equations ' : the two-factor theorems, he contends, are not only free from the ' artificiality of a multiple-factor technique such as your own and Dewar's, but also fit the psychological facts more accurately.' [1]

[1] I quote from a covering letter that he was good enough to send me with the draft of his first note to *Nature*, CXXXVI, p. 297, where he draws attention to the possibility of using these particular theorems for correlations between persons as well as between tests. In regard to the ' multiple-factor equations ' there seems to have been some confusion. Stephenson and others apparently supposed that these were put forward as a novel technique suited *only* to correlations between persons. As he wrote in the same letter : " You deserve great credit for trying to develop a new technique ; but to me it seems essential to use two-factor theorems." On the contrary, as will be seen in a moment, my whole practice has been based on the assumption that *no* ' new technique ' was required. Actually, in ' correlating persons ' both here and in America, two-factor theorems had already been used, where *a single general factor alone* was extracted, and (in a few researches) for *secondary* group-factors, where the persons correlated fell into discontinuous groups. In the draft memorandum to which Stephenson alludes, the multiple-factor method was preferred because it lent itself best to theoretical work in matrix

Incorporating the more important points brought forward in later publications, we may perhaps summarize these and other distinctive features of Q-technique as follows. To discuss them in any detail is hardly fair or possible, until we have seen more plainly how they actually work out in practice: my comments will therefore be relegated to the footnotes.

1. Both P-technique and R-technique, ' like all developments of psychometry up to the present day,' have concerned themselves solely with the study of individuals. But Q-technique " is in no way concerned with individual differences " but with the " deeper aspects of personality." Thus, the old R-technique sufficed as a basis for *applied* psychology; but Q-technique forms an entirely new branch of psychometry, which will serve as a basis for *general* psychology as distinct from applied, i.e. for theoretical psychology " as distinct from educational or vocational work in the field." ([138], p. 273).[1]

notation, to the more general proofs available with a matrix notation, and in particular to demonstrating the identity of the factors obtained by the two alternative approaches. The two-factor method was regarded, not as wholly inappropriate or invalid, but as a simplified procedure giving a first approximation only.

[1] On this point he is strongly opposed to the views that had previously been expressed. In my own *Memorandum* I had stated that " generally, the correlation of tests or traits leads to an analysis of the human mind in the abstract; and the correlation of testees leads to an analysis of the concrete human population," i.e. to the study of individuals and their grouping into types ([93], p. 253). This allocation of the two fields Stephenson would reverse. Nevertheless, ' Spearman's work with R-technique ' (at least as I understand it, and, I fancy, in the opinion of its author)—his studies of ' the nature of intelligence,' of ' the principles of cognition,' and of ' the abilities of man '—all this was surely primarily a contribution in the first instance to *general* psychology rather than applied: indeed, factor-measurements for *individuals* were, as a matter of fact, rarely calculated by the theoretical investigators who followed the so-called R-technique—e.g. Spearman, Stephenson himself, and others working in the same laboratory. On the other hand, for those of us whose daily business in the schools and school clinics was essentially concerned with ' individual differences,' correlating of persons naturally arose as a necessary device for " educational and vocational work in the field " as contrasted with the experimental work of the academic investigator. I do not, of course, deny that Q-technique or P-technique may be fruitfully applied to the problems of the theoretical psychologist as well:

2. The measurements with which Q-technique is concerned must be expressed in terms of an entirely new unit, namely, '*significance*.' This "concept of significance is peculiar to Q-technique and new to systematic psychometry. . . . No better example could be afforded of the difference between Q-technique and R-technique" ([136], p. 233 *f*). One trait is said to be more significant than another if it is "more representative or characteristic of a personality as an indivisible whole." Both P-technique and R-technique worked with isolated traits, which were measured, not by reference to the person who possesses them, but in terms of artificial norms which "at bottom are unscientific and unsound" ([92], p. 293): "the new unit of 'significance' will enable us to quantify qualities without tearing them from their immediate context." "R-technique took the person to pieces; only Q-technique can put him together again" ([98], p. 365, [96], p. 202).[1]

3. R-technique studied the *trait as variable*, with the persons as a normally distributed population; Q-technique studies the *person as variable*, with trait-measurements as a normally distributed population. Stephenson thus "draws a sharp distinction between statistical variables and statis-

I should rather argue that the contrast between the fields to which the two methods are appropriate cannot be pushed so far as Stephenson maintains, since when we correlate persons the factor-saturations obtained for the individuals tested are virtually the same as the factor-measurements obtained for the same individuals when we correlate traits.

[1] Once again I doubt whether the contrast is as radical as is assumed. 'Significance' seems to me merely a matter of weighting: if 'anger' is more 'representative' or 'characteristic' of a given person than 'fear,' that means we must give it a larger weighting in specifying his temperamental pattern. Moreover, in a research, dealing with say 60 traits and 200 persons (like Thurstone's recent study), factor-patterns characteristic of the 'whole person' can in theory be deduced just as well from correlations between traits as between persons. Actually, in giving instructions to his observers who are to rank each picture postcard for its significance to themselves and to grade temperamental traits for their significance to the patient, Stephenson uses almost exactly the same phrasing as the rest of us have done: his observers are asked to say which postcard they "liked best," or which mood or emotion is "most characteristic of the patient for its duration, prevalence, and its intensity": (these latter being the same criteria as I myself had used: "intensity, frequency, duration, and after-effects" [30], p. 694).

tical populations, which Burt appears to confuse" ([138], p. 275, [136], p. 20).[1]

4. R-technique has hitherto been concerned mainly with '*fractional*' and '*universal*' factors. The factors for which Q-technique is more particularly needed are '*non-fractional*' and, as a rule, *non-universal.* A non-fractional factor is defined as one that "refers to the whole communality of the variable" (here, of course, the 'person': [138], pp. 272–3). In more familiar terminology, this means that Q-technique, like Thurstone's method of 'simple structure,' will, as a rule, avoid extracting any 'positive general factor,' common to every correlated variable (i.e. common to all the persons); instead it will seek limited 'group-factors' by analysing certain clusters only (after the correlation table has first been 'suitably organized'), and will aim at 'oblique' factors rather than 'orthogonal' ([98], p. 357). It follows that *two*-factor analysis, as distinct from *multiple*-factor analysis by a general-factor method, must be employed.[2]

[1] Admittedly I do not draw a "*sharp* distinction between them," or rather, I regard the distinction as merely a formal distinction not a material distinction: cf. above, p. 180. It may be true to say that the difference is (as Stephenson puts it) "conveniently overlooked" when I use covariances; but that is not true when I use correlations, which alone are envisaged by Stephenson. Indeed, my list of the conditions which should normally be observed in *correlating* persons (as distinct from covariating) included the very points on which he now insists; and I have always stressed the peculiar difficulties involved in treating a selection of tests like a sample from a population ([101], p. 68).

[2] The 'clusters' thus obtained would seem to be virtually identical with the square diagonal 'submatrices,' analysed by the usual simple summation formula, in the so-called 'group-factor' method ('method a'). The identification seems justified by the fact that, where he extends his analysis into the oblong submatrices, i.e. those containing low or negative correlations produced by the overlapping of partly antagonistic group-factors (e.g. [136], p. 243), he himself adopts my formula ([116], p. 355, eq. viii). The chief difference is that I should nearly always begin by eliminating any general factor first: and the chief weakness is that Stephenson has apparently no criterion for his rearrangement of the correlations other than simple inspection of the correlations or persons correlated. Thus, in [98], pp. 364–5, unless he had previously selected persons *already known* to belong to the manic-depressive and schizophrenic 'types,' he would probably have made his calculations very differently; and in other tables, as Davies has shown [130], the divergences on which the clustering is based are often indistinguishable from the fluctuations due to sampling.

5. As a result, " Q-technique opens out an entirely new field, namely, typology." Types are defined as " any persons who satisfy the conditions for a non-fractional factor in Q-technique." And whereas R-technique was confined solely to ' *abilities* ' and ' *tendencies*,' Q-technique turns away from these artificial concepts and deals with the analysis of ' *types* '—a form of analysis (as Stephenson now holds) that could not be undertaken by the older methods.

Alleged Differences in the Logical Nature of the Techniques. —The whole theory is evidently full of interesting and original suggestions. Here, however, the question that we have to decide is not the validity, utility, or relative merits of P-, Q-, and R-technique, but the logical nature of the methods and the logical status of the factors that they yield. Does the device of correlating persons, or do the conditions under which it appears legitimate and fruitful, involve principles that are logically different from the procedures we have so far reviewed ?

My own opinion is decidedly that it does not. Some slight modifications are no doubt requisite (e.g. in regard to the minor but difficult problems of sampling error and the like). But for the rest I am wholly opposed to any sharp division of correlational data into two completely separate branches, each with its own stereotyped ' technique.' Always, and throughout every field of work, we are concerned in the last resort with relations between individual persons with their characteristic traits, on the one hand, and the various test-situations, on the other, and these test-situations themselves may equally arise out of the performances of other individual persons. Indeed, as often as not, it is a mere incident of the investigator's standpoint whether he regards a particular set of measurements as describing a ' person ' or as describing a ' test.' [1]

[1] Stephenson objects strongly to this view, protesting that the same row of figures cannot ' change chameleon-like ' from a ' population ' into a ' variable ' ([136], p. 21). However, his original exposition of Q-technique was based on this very assumption ([98], p. 354). Moreover, in some cases the material contrast between the rows and columns of the initial matrix may vanish almost entirely. A curious example arises in paired comparison. In contrasting the reliability or ' objectivity ' of judgments on pictures and on weight differences respectively (somewhat along the lines of Lyman Wells),

Our methods of analysis must therefore be elastic and eclectic, adapted specifically to the requirements of each particular research. Whether we use a multiple-factor or two-factor procedure, seek ' non-fractional ' group-factors or ' fractional ' common factors, trust to ' simple summation ' or prefer the ' method of least squares '—these are questions to be settled afresh for the problem under investigation, and not by *a priori*, cast-iron principles of Procrustean universality. In short, I hold, as stated in my original *Memorandum*, " that in principle, though not perhaps in detail, the statistical technique which has been worked out for factorizing the results of a test is equally valid for factorizing the operations of a person " ([93], p. 275). For this reason, in summarizing the available factor theorems, I described them as applicable, without essential change, *either* to correlations between tests *or* to correlations between persons.

That is precisely the view that Stephenson would controvert. " All earlier investigations based on correlating persons or extracting person-factors come far short of a general psychometric technique, avowedly defined as such, for the plain reason that it has refused to break with conceptions that are part and parcel of the old R-technique. Burt . . . still retains the standpoint of earlier psychometry ; his work on P-technique, therefore, appeared to Stephenson to have entirely missed the essential possibilities implicit in the idea of correlating persons " ([138], p. 276). Stephenson thus insists that his own technique

Eysenck has used paired comparison instead of ordinary ranking, and a correlational technique instead of studying mean variations (cf. Guilford, ' The Method of Paired Comparison as a Psychometric Method,' *Psych. Rev.*, XXXV, p. 494 f.) ; in such a case, if we take the usual table, whether we start by correlating rows or columns, the results will be the same, because the initial matrix is itself essentially symmetrical, and the variables for both are the same series of ' tests,' in the one case serving as ' standard ' or ' judge,' in the other case as ' variable ' or item ' judged.' On the other hand, where we have a number of psychologists assessing a number of children, the variables in both cases are ' persons.' Hence, as I have argued elsewhere, " the convenient antithesis between test and person is inexact and at times misleading : what appears as a test in one investigation may be treated as a person in another, and *vice versa*. The real antithesis is between comparing rows and comparing columns in one and the same matrix " ([101], p. 67).

is " methodologically as different from the older techniques as chalk from cheese."

But are the alleged differences so insuperable as he avers ? Do they really oblige us to " break with the old R-technique " ?

First, let us be clear where precisely the differences are said to lie. Too often it has been supposed that the differences referred to are not logical or psychological[1] but merely mathematical or statistical. Yet the two-factor theorems by which the calculations are to be carried out are not in themselves new theorems, nor even ' inversions ' of the old. Both the formulæ and the proofs are the same as in the case of R- and P-technique ; all that is different is the interpretation of the symbols, or alternatively the symbols themselves. And, of course, these formulæ or their equivalent had often been used before to factorize correlations between persons.[2]

As Stephenson himself repeats, it is not a new arithmetical procedure, but a ' new methodology, opening up a new field of psychology,' that he is anxious to advocate. The fundamental conceptions are different. Thus, for him ' typology ' requires us to envisage a ' fresh *kind* ' of factor, related to the traits on which it is based ' in a statistical and no longer in a mathematical fashion.' And this in turn introduces new logical principles into our demonstrations ([136], p. 280).

Alleged Differences in the Nature of the Factors.—It would take us too far afield to discuss these arguments in detail here. Perhaps the best criterion will be to look at the

[1] This is not Stephenson's own view. " In Burt's work on person-factors there is statistical novelty, but it is psychologically unimportant. In Q-technique there is nothing *statistically* novel. On the other hand, on the *psychological* side Q-technique opens up an entirely new field " ([138], p. 280, his italics).

[2] All the equations deduced in the papers on Q-technique ([97], [98]) were included in my earlier memorandum [93] as applicable to correlations between persons, though admittedly they were there depicted as narrower and less useful than the more general formulæ, based on the method of weighted summation. In the worked example ([93], p. 294) the saturation coefficients were calculated for comparison by both methods ; and it was shown that the differences were all less than ·005.

concrete nature of the factors themselves. In doing so, we must carefully distinguish between the first or positive 'general factor' and the secondary 'type-factors.' It is with the latter that Q-technique is primarily concerned. Accordingly, let us glance at the type-factors first.

1. *The Type-factors.*—"Q-technique," we are told, "is pre-eminently a method of isolating types. . . . In the past, type psychology could not be subjected to factor-analysis" ([98], p. 366). Now, as I shall endeavour to show later on, in virtue of the principle of reciprocity, the type-factors obtained by correlating persons are, in theory at any rate, virtually identical with those obtained by correlating traits. But we need not invoke a disputed principle; let us turn to actual results. A comparison of Stephenson's conclusions with those of previous investigators[1] would, I think, show that all the type-factors that he discovers are analogous to those that have been discovered by R-technique and by P-technique. Take, for example, his very first experiment [92] in which he proposed to test artistic appreciation or taste for the same kind of objects as Bulley, Pelling, and myself had used ("colours, vases, cloth materials, furniture, pictures, literature, and so forth"—with the addition of "perfumes" used by Beebe-Center ([92], p. 23). His two-factor procedure leads to two group-fac-

[1] Unfortunately many of the earlier efforts with so-called P-technique remain buried in degree and diploma theses; even when the final conclusions have been published, the cost of printing tables of correlations and saturations for a hundred or more persons has apparently been prohibitive. Davies, however, has brought together a summary of the chief problems attacked in this country and America; her review appears to include all except a few tentative experiments by junior workers ([130], cf. [101]). On the issue raised in the text, fuller evidence will no doubt be shortly available in the further results of research-students who have recently been working jointly under Dr. Stephenson and myself: as soon as we found that we differed so widely over the uses of correlating persons, a programme of work was drawn up covering the main fields in which divergences of opinion had arisen, with a view to setting investigators to examine each particular problem simultaneously by both procedures—by his method and by my own. Some of the preliminary results are quoted in Stephenson's articles; and the further data so far available—those of I. Cohen, P. C. Hu, M. Hill, and E. Knowles—would appear to strengthen the view that the discrepancies are more apparent than real.

tors, dividing his sample population of twenty into two groups: the first factor characterizing those persons who are "the more artistic and tasteful in general"; the second, the remaining persons who are "frankly inartistic." But since 'factor II' is admittedly the 'obverse' of 'factor I,' it is plain that the division into two 'types' is really effected by a single bipolar factor, namely, what in our experiment had been termed the 'general factor' for 'artistic taste.' [1]

When we turn from æsthetic to temperamental types, it would seem that Stephenson's new factors, as obtained by Q-technique, are very similar to those he had previously obtained by R-technique. In the same issue of *Character and Personality* there appears a short series of articles by his collaborators and himself on 'The Factor Theory in the Field of Personality.' The "questions to be studied" are much the same as those dealt with by Q-technique: "(1) what are the unitary or group-factors to be assessed in considering personality? (2) how are they to be measured? and (3) how are they related to the total personality?" The results at first sight appear almost identical with those I myself had obtained by analysing correlations between emotional tendencies. "One unitary (i.e. general) factor was . . . verified in the present research, . . . and a second factor revealed." The former appeared closely related to so-called tests of perseveration (p), but is ultimately identified with Webb's w (persistence or stability) rather than perseveration; the latter (f) is termed 'fluency,' and appears as a bipolar factor. "Prof. Burt," it is added, "points out two groups of emotional tendencies, which he terms sthenic (assertive or unrepressed) and asthenic (or inhibited), and these terms would appear to fit the fluency factor admirably." These factors are also said to account for the current classification of temperamental types (e.g. 'high f' is correlated with the 'explosive' or unrepressed

[1] The figures he prints are illustrative only; and no evidence is offered for or against the more specific 'types of artistic appreciation' which previous work had suggested. Stephenson's next experiment, however (with picture postcards [97]), seems to indicate a bipolar factor yielding more specialized types analogous to our own, viz. the 'classical' (with preferences for what is formal) and 'romantic' (with preferences for the sensational and ornate).

type, ' low *f* ' with the ' obstructed ' or repressed). To illustrate his views Stephenson discusses a typical extravert and a typical introvert in terms of these factors ; and his analysis of clinical cases is said to show " a correspondence with clinical types when patients are examined in terms of these factors and of their relations " (pp. 37, 43, cf. 51). All these conclusions are based on correlations between traits. Strangely enough, nothing whatever is said in any of the papers about correlating persons, although in the article a few pages back, describing that method, " correlating persons instead of tests " was said to be the " essential tool to check the theories of type psychologists."

In a symposium on the same subject [91], all the contributors seem to have assumed that R-technique was entirely adequate. Hitherto, it is stated, the work of the ' Spearman school ' has been chiefly confined to the " field of individual differences " as distinct from " general or ' type ' psychology." In the latter field " it is easy to work to ' types,' such as introvert or extravert, but difficult to measure the degrees ; yet . . . factors of the Spearman kind are the *only means* whereby these quantitative suggestions can be justified and regularized" (p. 102, my italics). Such factors, deduced by correlating tests or traits, form the best ' classificatory device ' for use in psychiatry : " the whole individual can be described in their terms." Stephenson's first paper closes with a section on ' The Future of Factor Studies in Psychiatry ' ; yet even here there is not a word about the utility of correlating persons.

Dr. Stephenson's conversion to the alternative procedure, therefore, seems, not only sudden, but even startling in its thoroughness. For in his next publication he gives an analysis of clinical types among patients from the same mental hospital, carried out this time by Q-technique, and now assures us that this is the sole method for " isolating types and measuring persons for their approximation to such types " (the schizophrene and the manic-depressive among abnormals, introversion and extraversion among the normal). " It has never before been possible," we are told, " to measure for types " ([98], pp. 357, 366). Surely,

this not only contradicts, but does great injustice to, the previous researches of Stephenson himself and his collaborators on this very subject. Is not the true conclusion one which will reconcile both approaches, and regard them as alternative and more or less equivalent ways of reaching the same result ?

2. *The General Factors.*—The problem of temperamental types and factors I shall discuss more fully, in the light of further data, in Part III. Meanwhile, in two recent *Contributions to the Theory of Mental Testing* [136], Stephenson has returned to the attack from a somewhat different angle. In these papers he describes a new use for Q-technique, namely, ' measuring abilities as correlation coefficients.' Here, evidently, we are no longer concerned with type-factors, but with general or universal factors such as were formerly investigated by R-technique. He begins by stating that " work on the correlation of persons has been conceived by Burt as merely [1] an alternative procedure to that of correlating tests or traits, it being a matter of convenience whether we correlate persons or tests." In contrast to this conception, he proposes " to demonstrate that Q-technique has its own special problems, and that its assumptions, principles, and objectives are peculiar to itself, and independent of present-day testing-theory " ; and his argument is illustrated in the concrete by an experiment with a new and ingenious performance test for measuring intelligence by a sample of tasks, drawn, as it were, from a homogeneous ' population ' of possible test-material.

To support his version of my view he refers to a couple of my recent articles ([101], [114]). In the passages cited, however, I expressly mentioned ([101], pp. 66, 72, 81 ; [114], pp. 166–7, 181, 187) that I was dealing solely with bipolar (or type-) factors : the general factor for persons I had already dealt with in a previous publication ; and

[1] And again on p. 34 he repeats " accordingly to Burt it is *merely* a matter of convenience whether we correlate person or traits." The error lies in the insertion of the adverb ' merely.' On the page following the one from which he quotes, I went on to consider the cases in which ' we find ourselves *compelled* to correlate persons with one another or with an ideal person " (pp. 62–3).

there it was pointed out that this factor could *only* be obtained by correlating persons, and that it enabled us to "measure the efficiency of a *person* by a coefficient of correlation, just as with ordinary factor-analysis it has long been customary to measure the efficiency of a *test* by such a coefficient" ([93], p. 275).

But that is precisely what Stephenson's new test sets out to do. Each examinee's ranking of 20 test-items is to be correlated with a standard ranking which represents that of the ideal expert; and "the correlations so obtained are a direct measure of his ability." The test itself, he explains, is "like any other of a performance type, except that it embraces a more varied sample of intelligent activities"; the correlations are "proportional approximately to ability as measured in the more usual way." Now the factorist's "more usual way" of measuring ability is, of course, to correlate tests, and take a weighted average of those tests of 'intelligent activities' that yield the highest correlations with the general factor. The reader, therefore, is not a little astonished to discover that Stephenson has chosen, as a typical problem in which it is "necessary and sufficient to correlate *persons*," the very problem which has always been studied by correlating *tests*—namely, the measurement of general ability or intelligence. Moreover, we are told that "it is not for a moment suggested that we should in practice measure abilities as correlation coefficients" (i.e. as correlations between persons): so that it turns out, after all, to be largely "a matter of practical convenience" whether we proceed by Q-technique or the old-fashioned R-technique.

At the same time, I fully agree that there may be "conditions under which it is necessary and sufficient for us to correlate persons rather than tests." Indeed, elsewhere I have attempted to indicate both the 'necessity' and the 'sufficiency' in several different fields of inquiry.

If the reader will reflect once again on the illustrative problems already cited, he will see quite plainly how the necessity arises, and will glean some notion of the general type of inquiry for which this particular approach (as I understand it) is more especially suited. The clearest instance occurs in studying the 'reliability' (i.e. the

mutual agreement) of examiners or observers grading or judging a series of individual items. In evaluating the Binet-Simon tests, for example, I tried to show that there was " a central factor, presumably the ideal order," underlying all the individual orders given by the different examiners; such a 'central factor' could not possibly have been demonstrated, had I started by correlating tests.[1] Again, in investigations on the degree of agreement among school and university examiners, we endeavoured to show that, as a rule, there was a "common factor, namely, the ideal order of merit," influencing the marking of each one; this again could never have been established had we correlated the marks by scripts instead.[2] Somewhat closer to Stephenson's 'new method of testing' are the various ranking tests we have tentatively sought to construct for moral, æsthetic, humorous, and logical judgment, in which the examinee is required to arrange series of moral offences, poetic extracts, jokes, comic pictures, logical absurdities, musical pieces for a concert programme, etc., etc., in order, and his grading is correlated with a weighted average.[3] Without correlating persons, for instance, it would have been impossible to discover " an apparently objective element in judgments on artistic taste . . . some one general underlying factor influencing the judgments of all "; moreover, a practicable method of measurement was at the same time described: " by correlating the order furnished by the individual tested with the order furnished by art-critics, taken as a standard, we could," it was stated, "obtain a coefficient which might be a mark or measure " for that individual's artistic capacity or " taste."[4] The same principle was tentatively applied to the familiar absurdities test to obtain assessments of intelligence for vocational guidance with supernormal adults: the examinees were given a series of short reasoned statements of varying validity, and asked, not to point out the absurdities, but to arrange the arguments in order, putting the most cogent first and the most absurd last: the 'intuitive logical judgement' of each was then assessed by correlating his order with

[1] *Mental and Scholastic Tests* (1921), p. 136 and Table 137.

[2] *Marks of Examiners* (1936), pp. 275, 292; also [29] and [134].

[3] Cf. *The Young Delinquent* (1925), pp. 404–5, and theses by Moore, Fraser, Wood, Williams, Dewar, Wing, and others, referred to below.

[4] I.e. what is technically called his 'saturation coefficient' for the factor. *How the Mind Works* (1933), p. 292; cf. also [118], p. 32 and Table I. As will be gathered from the preliminary account, several hundred adults and children, of different ages and sexes, were measured by Miss Pelling and myself in this way; and the correlations between the measurements and the teachers' judgments (particularly at the L.C.C. Art Schools) proved encouragingly high (·66 to ·79).

the best order.[1] Similar methods have been used by Schonfeld for assessing and factorizing the 'sense of humour,' by Dewar for pictorial appreciation, by Eysenck for other forms of visual art, by Wright and Wing for musical ability, by Williams and Wood for literary appreciation. In every one of these investigations it seemed necessary to correlate persons rather than tests or traits.

What, then, in general terms, are the " conditions under which it is necessary " to adopt this particular procedure ? Since Dr. Stephenson thinks that I have overlooked them, may I briefly refer to earlier passages (some of them in the two articles he himself cites, e.g. [101], pp. 61–6, [114], pp. 176–7 ; cf. also [130], pp. 416 f.) in which I attempted to summarize them, both in general and in detail ? A comparison of the common features in the inquiries just enumerated will indicate the general nature of the problems and purposes for which, as I argued, we seem " compelled " to adopt this alternative procedure.

Grounds for Correlating by Persons.—Broadly speaking, we correlate persons rather than tests when we are concerned with the complex resemblances between total personalities (or aspects of those personalities) rather than with the more limited resemblances between particular traits or their tests. And, whenever we desire to investigate what I have called the ' general factor for persons,' these are the correlations that it is *necessary* (and not merely convenient) to factorize. In psychological work this general factor may be a cognitive, affective, or conative factor, a combination of the three, or a sheer artefact ; it may be a factor of ' reliability,' of ' preference,' or ' judgement,' or of efficiency measured by any of these three, or a mere ' halo-effect ' ; in short, it may be any abstract characteristic which is ' subjective ' [2] in the sense that it

[1] ' A Judgement Test for Measuring Intelligence,' *Mental Welfare*, XX, pp. 45–48. Here the correlation with independent assessments of the examinees was not so high as Stephenson's (·67 to ·73).

[2] Stephenson takes exception to my " limiting inverted factor-technique to problems of ' subjective ' judgement." He prefers, as we have seen, to use the more general term ' significance.' Elsewhere he says " only the psychologist can decide upon matters of significance ; but that by no means implies that appraisal for significance is subjective " : for (i) the psychologist

sums up a distribution or pattern within one and the same person, and 'unipolar'[1] in the sense that its saturation coefficients in the different persons are always positive. In investigations with cognitive tests, the 'general factor for tests' is the factor obtained by averaging all the tests for each person; and in the same way the 'general factor for persons' is the factor obtained by averaging[2] all the persons for each test. Thus, if we are chiefly interested in comparing the tests, we can call this 'general factor for persons' the relative ease or *difficulty of the tests* (it will usually vary with their complexity and level); if we are chiefly interested in comparing the abilities, we can call it the *relative strength of the abilities* in the whole population tested (repeating eight numbers is a harder test than repeating four numbers because the average person's ability to do the latter is greater than his ability to do the former). In investigations on emotional traits—e.g. affective preferences or conative tendencies, the 'general factor for persons' will similarly be the relative strength of those preferences or tendencies in the general population. And in investigations of almost every kind, since the resemblances studied are complex, the 'general factor for persons' will (to put it briefly) describe a characteristic *pattern* of measurements that distinguishes the universe of individuals for whom the factor is 'general' or *generic*: e.g. in human

is a technical expert, and (ii) his appraisal can be checked against those of others ([136], p. 236). Nevertheless, these features by themselves do not convert an individual grading into an objective measurement. On the other hand, I admit that the underlying general factor (if any) is 'objective' in that it is, by hypothesis, independent of any *individual* judgment. Indeed, one of the purposes of my research was to show that artistic taste depends largely on an objective factor in the same sense that weight-discrimination does, though not to the same degree ([75], pp. 289, 294).

[1] If I am right in regarding the table of positive correlations that is commonly found in correlating tests as really a single quadrant of a larger bipolar table (see above, p. 163), then the distinction between general and bipolar factors would also disappear. This, however, is too speculative a point to bring forward here, and would only confuse the issue.

[2] The notion that a set of factor-measurements may be obtained by *averaging* (weighted or unweighted) is another point in my theory to which Stephenson emphatically takes exception (e.g. [98], p. 357 and ref.). However, this is a point which I hope to demonstrate in the last part of this book.

psychology, the human genus; in group psychology, the particular group selected for assessment (cf. above, p. 113).

Curiously enough, in his earliest discussions of correlations between persons, this was the line of work that Stephenson most vigorously criticized. As we have seen, his chief criticism was that we interpreted our results in terms of "abilities" instead of "types"—in terms of "universal factors" common to all the persons correlated instead of "type-factors" limited to particular groups. It was even said that the method had led us to "deny the existence of psychological types" ([92], p. 295). Thus, when he now comes forward to show that "correlations between persons may themselves be a direct measure of *ability*," he is not opposing my original view, but reverting to it[1]: for his new application of Q-technique has nothing to do with 'typology,' but solely with the measurement of the general factor of intelligence.

It is not difficult to see how the double misunderstanding has arisen. The 'two-factor theorems' on which Stephen-

[1] I am tempted to say that our procedure seems to fulfil Stephenson's present requirements much more completely than his own. His article in its opening paragraph claims to be an illustration of "Q-technique, that is, the factor-study of persons as variables" and to be "concerned with the general theory of factor-analysis." Actually, however, there is neither factor-analysis nor any study of factors as such. His ideal order is not obtained by factorization of correlations, but is directly deduced from "the logical order of decretion on which the test was constructed" (p. 32). Although, therefore, he sets out to show the unique conclusions to be attained by analysing inter-correlations between persons, in the end he does not correlate *persons* at all! What he correlates is one person's arrangement with an *objective* standard, much as Binet compares a child's arrangement of given weights with the order of heaviness assigned by the balance. The 'standard person,' who is set up to serve as a reference value "much as length is measured in terms of a yard," though several times referred to in the introductory argument, is never actually derived or used, and seems wholly unnecessary to his procedure. In our own procedure the "standard personality"—the hypothetical 'typical individual who serves as a common point of reference'—is a kind of weighted average, specified by the 'factor-measurements for tests.'

At the same time I do not wish for one moment to decry the value of Stephenson's highly ingenious test. If I understand his figures correctly, it would seem to yield a decidedly closer agreement with independent criteria of intelligence than my own, and at the same time to be freer from the complicating influence of verbal ability.

son has inflexibly insisted do not recognize the distinction which the 'multiple-factor equations' preserve between the first 'universal' or 'general' factor (with positive saturations for every correlated variable) and the secondary 'group' or 'bipolar' factors, specifying groups or types. Hence, for him, if a type-factor is not specific, it must be general. Consequently, with his procedure my secondary *type*-factors became *general* factors (in practice nearly always calculated for certain sub-matrices only); and conversely, he has maintained, at any rate in his earlier papers, that my general factors, when procured by correlating persons, "could only indicate types," and therefore should really have been type-factors.[1] Thus he applies (rightly on his premisses, wrongly on mine) everything that I have said about secondary factors to the general factors as well, and *vice versa*.

Conclusions and Corollaries.—In conclusion, let me briefly indicate what I take to be the logical nature of the difference between us. At bottom, I think, the difference is due, like so many controversies in factor-analysis, to a confusion between two logical standpoints—namely, the analysis of 'intension' with the analysis of 'extension.' In correlating traits we are (in the first stage) comparing the attributes or traits that define certain classes; in correlating persons

[1] This appeared still more plausible because, as explained above, in repeating our experiments, he rejected our relatively heterogeneous test-material, and substituted pictures, etc., that were far more homogeneous: but such a modification must inevitably tend to eliminate the general factor with which we were primarily concerned, and to throw into relief our 'secondary type-factors.' This equality (actual or assumed) in the difficulty of the various tests explains a peculiarity that marks so many of the tables obtained by correlating persons. Thomson suggests that "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" ([132], p. 211). But the experimental difference arises only when the investigator works with tests or traits that have, or are assumed to have, the same average level for the different persons: thus in the investigations to which Thomson alludes the 'general factor for persons' had virtually been partialed out beforehand or at least reduced in magnitude. Under these conditions, as I have elsewhere explained, a "bipolar" table of coefficients is "apt to be the first to appear when correlating persons" ([114], p. 167, cf. [130], p. 419).

we are (in the first stage) comparing the members of those classes. (By the first stage I mean the stage which carries the factorist as far as computing the factor-saturations.) Thus, in the former case, we take an intensional quality that can be predicated of the various persons, such as ability to read, and subdivide it into a generic predicate, say intelligence or cognition generally, and a specific predicate, say verbal capacity. In the latter case we take an extensional class, namely, the children of whom such abilities are predicated, and subdivide it into the more intelligent sub-class and the less intelligent sub-class, and each sub-class again into the verbal and the non-verbal type. To declare that by correlating persons we can discover types that cannot conceivably be discovered by correlating traits is like declaring that there are classes for which no defining predicates (i.e. no trait-combinations) could be possibly found.

The supporters of Q-technique rarely, if ever, go on to the second stage and calculate factor-measurements, whether they are using Q- or R-technique. When we proceed from factor-saturations to factor-measurements, a transition is made from intension to extension in correlating traits and from extension to intension in correlating persons. Once again I find it difficult to imagine that, by Q-technique, I can obtain a factor-saturation describing a set of persons as a type, and yet, in principle, be entirely unable to reach the same figure by correlating the relevant traits: it is like saying the name JONES can only be found in the telephone directory by hunting for the word as a whole, and that if I look first for J, then for the O's under J, then for N, E, and S, I shall never find Jones. The equivalence of the two procedures, however, is a matter that I shall defend more fully when I deal with the so-called 'reciprocity principle' (see chap. xi).

It is perhaps too early to come to any final decision in regard to the status of 'inverted-factor' techniques. But, as I see the matter, Stephenson's criticisms of my own procedure, and certainly my criticisms of Stephenson's, relate to comparatively minor issues. On essentials we seem in pretty close agreement. I do not deny that, as regards statistical significance and the effects of selection, and more

particularly the psychological interpretation of the factors themselves, many new problems arise, some as yet unsolved ; but that does not of itself commit us to a ' completely new methodology.' My provisional conclusion therefore is that, although there may be material differences between the factors obtained by the different ' techniques,' formally the newer applications involve no fresh logical principles, and lead to factors in no way radically different from those obtained by the older.

Nevertheless, as I have from the outset insisted, when we turn from the older applications to the newer, fresh light is unquestionably shed both on the logical nature of factor-analysis and on the intrinsic nature of the factors themselves. Though that light may reveal no novel features hitherto unsuspected, it throws many of the previous conclusions into sharper relief. Two important inferences that are thus very clearly confirmed deserve amplifying at somewhat greater length.

Factors as Principles for Classifying Tests.—In the commoner applications of factor-analysis, depending on the correlation of tests, the investigator, as we have seen, usually identifies his factors with abilities in the minds of the persons tested : the ' factor-saturations for tests ' are taken to state how far each test depends upon, or is saturated with, the particular ability specified, i.e. how far the results of the empirical test would resemble or correlate with the results of a perfect test for that ability. The ' factor-measurements for persons,' which are then deduced, indicate the results that we should expect such a perfect test to yield, i.e. they provide quantitative gradings or classifications for the persons in respect of the hypothetical ability : thus, if a particular examinee has a large negative measurement for the ' general factor ' of intelligence, exceeding some borderline figure, he will for practical purposes be classified as ' mentally defective ' ; similarly, if he has a large positive measurement for the ' verbal factor,' he can be classified as belonging to the ' verbal type.'

Now, when we adopt the alternative approach and start by correlating persons, we reach similar ' factor-measurements ' for tests instead of for persons. By analogy we must

now regard the factor-measurements as grading or classifying, not the persons, but the tests. At the same time, since we have started with the same set of data, we must surely assume that the factors—i.e. the principles of classification—will still remain the same, whichever approach we follow. But, under appropriate conditions, as we have seen, the 'factor-measurements for tests' obtained by correlating persons are identical with the 'factor-saturations for tests' obtained by correlating tests. Thus, whether we correlate tests or persons, we must regard our factors *as being principles for classifying tests or traits quite as much as principles for classifying persons.*

This conclusion is perhaps most obvious when we think of the familiar tests of school attainments. Usually the factors that emerge are described as revealing certain elementary capacities in the mind—verbal ability, manual ability, arithmetical ability, general intelligence, and so forth; but they may equally be conceived as principles for classifying school subjects. Thus we may group together reading, spelling, English composition, as verbal subjects; handwriting, drawing, painting, woodwork, as manual subjects; arithmetic, geometry, algebra, and physical science as arithmetical or mathematical subjects: and each subject may, as it were, be cross-divided into levels of difficulty, corresponding with the level of intelligence possessed by the children who can cope with it.

Such an example shows very plainly how the factors deduced from any set of correlations depend primarily upon the way in which the tests correlated have been chosen: and it prompts the further inference that, if factor-analysis alone were sufficient to reveal 'abilities,' we should have just as much right to speak of the 'abilities' of our tests as of the 'abilities' of the persons tested.[1] The solution to the paradox is not far to seek: at bottom what we are really classifying are neither the minds, on the one hand, nor the tests or traits or school subjects, on the other, but the *relations* between the two.

The same holds good in investigations based on tests of a

[1] Beebe-Center, indeed, has actually used this phraseology; cf., however, [97], p. 349, and [130], p. 415.

laboratory type, or on assessments of character-qualities and traits of personality. Whipple's well-known manual, for example, presents all the more familiar mental tests classified according to the functions which they were presumably intended to measure—sensory discrimination, motor capacity, attention, memory, and the like. And the various questionnaires issued year after year for recording children's behaviour nearly always presuppose some tacit or explicit classification of the actions to be observed—some hypothetical scheme of 'propensities,' 'attitudes,' or 'unitary traits.' Evidently, if such classifications are to have any genuine value, they themselves require first to be substantiated. And in either case this can only be achieved by a comprehensive factorial analysis.

Factors specify Relations.—The view to which we have been led becomes almost unavoidable, when we reflect on the very different constructions that may be put on the results of correlating persons: here, as we have seen, investigators find it much more difficult to agree over what precisely they have been analysing. Correlations between traits or tests are almost universally assumed to yield an analysis of the mind into its component 'abilities': correlations between persons, on the other hand, may be interpreted in two seemingly inconsistent ways.

Consider once again the type of inquiry I have been describing—for example, one of the many experiments in which a number of persons are asked to rank a series of English compositions in order of merit, and the rankings are correlated by persons. What do the results reveal? Do they indicate the literary appreciation of the judges, or the literary merits of the compositions judged? If the compositions are children's essays, and the judges school examiners, then we shall probably regard the figures as grading or classifying the compositions judged. If, however, the judges are children and the compositions extracts from authors differing widely in literary skill, then we may use the figures to grade or classify the judges according to their powers of appreciation. The truth is that with both experiments what our measurements really express are in the first instance neither the qualities of the judges by

themselves, nor the qualities of the test stimuli by themselves, but the interaction between the two. Which of the two interacting qualities we treat as constants and which we treat as variables will depend upon the question we set out to study and upon the way the experiment has been planned. This conclusion, which seems more or less obvious when we are correlating persons, is, I should maintain, equally true when we are correlating tests. In either case, the psychologist's factors do not in themselves describe either persons stimulated or things used to stimulate them; they describe *the complex relations between the two.*

Principle of Reversibility.—To regard the mathematical analysis as concerned, not so much with the relata, as with the relations, will confer much the same advantages in psychology as in physics. It abolishes the need for fixed distinctions between 'determiner' and 'determinate,' between known 'ground' and unknown but predictable 'consequence,' or (to use more popular terminology) between 'causes' and 'effects.' In experimental psychology we talk of tests as stimuli which cause reactions; in factor-analysis we more commonly talk of abilities as "causes of individual differences."[1] I should prefer to regard each test-process, not as a direct and simple measurement of an isolated trait or an isolated individual, but as a statement of the point at which an equilibrium is reached between the external forces (the difficulty of the constituent test-items) and the internal forces (the mental activities of the individual tested). Thus, instead of rigidly distinguishing between variables that are given (cause-factors) and those which have to be deduced (effect-factors), as we do in applied or practical psychology, pure psychology should be content with a set of 'equations of condition,' which all the variables taken together would have to satisfy. The number of equations would then correspond to the number of variables that we *might* require to deduce or determine, without stating whether the deducible variables are 'abilities' or future 'performances' or possibly some combination of the two.

Accordingly, if we drop the usual antithesis between

[1] E.g. Thurstone, [84], p. 45.

tests or abilities as causes, and individual performances as their effects, then, as in physics, our equations should in general become reciprocal or reversible. Within the limits imposed by errors of measurement and sampling, we should be able to deduce trait-measurements from factor-measurements by the means of the same equations that are used to deduce factor-measurements from trait-measurements. And once again we shall be led to identify person-factors and test-factors ; for our conclusion implies that the factor-measurements derived from observations on the same sample of traits in the same or similar samples of the population should in principle be equivalent whether we proceed by grouping traits to correlate persons or by grouping persons to correlate traits. To suppose that the two sets of factors will completely change when we turn from one alternative procedure to the other would be rather like announcing that Boyle's law will hold good if you consider the effect of changing temperature upon the volume of a certain gas, but that a different formula will be required when you treat the change of volume as cause and the change of temperature as effect.

Cautions for Research.—The foregoing conclusions lead at once to several obvious corollaries of immediate importance for the research worker. More especially, we see at almost every point *how dependent the emerging factors will be, not only upon the initial choice of persons to be tested or observed, but also upon the initial choice of tests or traits.* The early factorists believed that the factors they were discovering were clear-cut, concrete entities, forming part of the well-defined structure of the mind : it scarcely mattered what tests were picked for correlation, the general factor and the group-factors (if any) would surely disclose themselves. And their later followers have further maintained that " no matter what sample of the population we take, our technique will enable us to classify them into types as distinct as primroses and buttercups." These two beliefs are more in keeping with the views of the scholastic logicians than with those of the present day. The first belief seems implicitly to take for granted the old traditional notion that the behaviour both of the mind and of material substances

is to be ascribed to a set of inherent, arbitrary, and independent powers; the second contention (which is put forward as a consequence of the first) implies a faith, like that which lasted from Aristotle to Linnæus, in the definiteness and stability of genera and species. And the student who reads some of the most recent publications comes away with ideas that are indistinguishable from the ancient doctrines of mental faculties and mental types in the crudest and most primitive form.

As regards the structure of the mind, the more cautious investigators, factorial as well as non-factorial, seem unanimous in agreeing that it is anything but fixed, sharply localized, and rigidly defined. Thomson, as we have seen, regards the mind as almost structureless; and few writers would nowadays explicitly assert that it is an aggregate of isolated powers. Similarly, with the exception of certain psychiatric writers and the supporters of 'Q-technique,' few modern psychologists, and hardly any factorists, would now maintain that the human population is divisible into a number of sharply distinguishable, mutually exclusive types, each characterized by some specialized ability, or by some specialized set of temperamental characteristics, which the others do not possess: in the few instances in which the statistical results exhibit some such discontinuous and multimodal distribution, the effect can usually be explained by unconscious selection in the initial data.

Those of us who have been directly engaged in practical work on school children (the field of work which is at once most freely accessible and most easily controlled) have always emphasized the changing nature of the factorial picture, as it is deformed by irrelevant or unintended selection. The more general factors, which appear so conspicuous when we are dealing with random samples or complete age-groups, seem almost to vanish when we apply the same tests to selected populations, such as pupils from secondary schools or from homogeneous forms or classes. On the other hand, the more limited group-factors, which are so hard to demonstrate in small random samples, often stand out in salient relief when we turn to a picked batch of scholarship children or of mental defectives.

The influence of selecting persons is now fairly well recognized.[1] The influence of selecting tests is more frequently overlooked. The technique of correlating persons demonstrates very clearly how the special selection of tests or traits can twist the whole factorial pattern. Looking back, such effects now seem sufficiently obvious, even in earlier work where tests rather than persons were correlated. The first investigators began by selecting simple sensory tests, and identified intelligence with general sensory discrimination. As they turned to more complex tests taken from higher mental levels, and again to traits of an emotional or conative rather than cognitive character, they first broadened the single general factor to cover the whole of the mind, and then limited it to make room for a second general factor of emotion or will. The stability of our factors, therefore, is only a relative stability. But, after all, that is true of most other abstract measurable concepts—the length of what we call a year, or the weight of what we call a pound of lead. In psychological work, no doubt, particularly until we know more of the genetic, physiological, and biochemical bases of the mind, the instability is likely to be even greater than it is in the physical sciences.

[1] It is still frequently disregarded in research theses; and largely accounts for the conflicting results obtained in early days, and for the conflicting theories that were consequently deduced. Rereading the descriptions and noting the figures for the variability, the student will continually perceive how the early advocates of *g* were prone to draw inferences from small but heterogeneous samples; on the other hand, those who discovered "no *g*, but only group-factors," were often arguing from experiments on relatively homogeneous classes or schools, where the basis of selection had already more or less equalized the *g* of the various members.

The influence of the choice of tests comes plainly to the fore when we deal with correlations between persons. Such correlations "can most safely be interpreted, if the tests (or traits) form a random sample of their total population"; and, as I have shown at length elsewhere, "on deducing the consequences of this condition, we are led to theorems closely resembling Thomson's sampling theorems, except that our propositions now refer, not to the fact that our tests are sampling the mind's neural 'bonds,' but to the fact that *the psychologist is sampling the tests or traits*" ([137], pp. 88–9). And in my view it is *the mode of sampling the 'tests,' not the 'bonds,'* that should be invoked to explain the hierarchical tendencies which correlations between tests so commonly display (*ibid.*, pp. 85–6).

Yet this only means that in defining his fundamental concepts the psychologist must, provisionally at any rate, admit a somewhat larger element of convention. We must specify our standard population and our standard set of tests. In the case of general intelligence, this has already been attempted, though not always with a clear consciousness of the reasons. There the idea of a typical sample of the population, with a typical standard deviation, and a typical test-scale, like the Binet-Simon series, has already been adopted for practical purposes. The same principle could be explicitly accepted for other traits and for theoretical work; and the best provisional conventions might thereupon be discussed and defined.

Let us agree, then, that mathematical analysis by itself, and apart from all other considerations and inquiries, cannot ever suffice to disclose what mental factors make up the human mind or what mental types are discoverable within the human race. But for the same reason, it follows that it is equally illogical to condemn a method of factor-analysis as invalid because its results in any particular cases are 'devoid of psychological significance.' If the results have no meaning, what stands convicted is, not the statistical technique, but the planning of the experiment, or more probably the lack of planning—the rash collection of data without prior regard for the needs of what from a mathematical standpoint is the best statistical procedure, and what from a psychological standpoint is the most relevant selection of measurable facts. Factor-analysis, as we have seen, is simply a form of averaging: whether the resulting averages have a plausible meaning or not depends on what we have decided to average. Neurological, biological, genetic, introspective evidence should determine both what traits and tests we include at the outset, and what concepts are permissible in constructing the final interpretation. If in the end our analysis hands us a meaningless answer, that is because we have asked a meaningless question.

CHAPTER VII

THE METAPHYSICAL STATUS OF MENTAL FACTORS

General Problem.—We have now examined the general nature of factor-analysis as a logical procedure—a procedure applied more specifically to psychological problems, but equally applicable to analogous problems in other complex sciences. We have also reviewed the chief factorial doctrines to which that procedure has given birth. It remains to consider the theoretical and practical bearing of the results at which we have arrived.

The wider implications of factorial work for theory have been strangely ignored. Not only does it raise issues of far-reaching importance to mental science and mental philosophy, but it also seems to indicate, or at any rate to limit, the nature of the possible replies. More particularly, the view we have been led to take of it should, I think, help us to supply a more decisive answer to the fundamental question raised or begged by the various uses to which factors have been put: namely, in what sense can the mind be said to contain or consist of factors? Or, to put it more specifically, how far are we justified in attributing a concrete reality and a causal efficacy to the factors we deduce?

The Definition of Factors.—Before embarking on these deeper issues, we ought first to make sure that we are all agreed on the meaning to be attached to the word 'factor.' So far I have refrained from defining the term, because I hold that any such definition should come at the end of the inquiry rather than at the beginning. Let us continue, therefore, to interpret it by saying 'factors are what factor-analysts seek and find'; and let us briefly examine their own descriptions.

It is curious that hardly any factorist has offered a formal definition. Thurstone explicitly defines 'traits,' 'abilities,'

'reference abilities,' 'primary traits,' and the like; but he does not define 'factors.' Warren, indeed, describes a factor as "a psychoneural element or determiner which is fundamental to all correlated abilities for the same individual"; but his description is apparently limited to factors of a particular type.[1] Spearman, to whose writings the term chiefly owes its currency, introduces them as 'organs' or 'fundamental functions' of the mind: the general factor he ultimately explains as a form of 'energy,' and the more specific factors as nervous mechanisms or 'engines.' This, he admits, is to interpret rather than to define. Nevertheless, whatever its ultimate nature, a factor, he insists, must at least be something more than a mere average or sum; and he offers a "proof that *g* and *s* exist." "If meaningful as opposed to statistical," he writes in a more recent paper, "a factor is taken to be one of the circumstances, facts, or influences which produce a result"[2]: in short, it is some kind of causative agency, even though the actual form of that agency is as yet unknown

[1] H. C. Warren, *Dictionary of Psychology*. Actually this is given as an explanation of a 'general factor,' according to the two-factor theory: under the word 'factor' no definition is given that would seem appropriate in factor-analysis.

[2] This broader definition is given in a recent reply to Thomson's protest against assuming that factors are 'entities or organs' (*Brit. J. Psych.*, XXIX, 1938, p. 184). The definition is evidently quoted from the *Concise Oxford Dictionary* (s.v.); but of the definitions there given I should prefer for factorial work the one that immediately precedes it: "(Math.) One of the components that make up a number or expression by multiplication." I would add that if the word 'component' had become current instead of 'factor,' there would have been far less temptation on the part of students and others to assume that 'factors' were concrete agencies, combining to produce a concrete 'product.' Much the same confusion is discernible in the history of physics. Current textbooks still talk of 'a beam of white light as a mixture containing red, green, and blue light, etc.'; and it is not so many decades since physicists were still disputing whether Newton's experiment on spectral analysis had or had not demonstrated by experiment the real existence of red light, etc., in the white. White light, of course, has no periodicity; but any disturbance, however irregular, may be analysed mathematically into a sum of Fourier components. Nowadays, I imagine, no physicist would insist on arguing that the component waves 'really existed' in the white and that it was wrong to regard them as mere 'mathematical artefacts.'

([56], pp. 61 f.). Similarly, Alexander declares: "We do not seek functional entities (i.e. entities that are mere mathematical functions), but psychological entities; our purpose is to resolve abilities into their true psychological factors" ([82], p. 3). And Stephenson: "A factor must be a real and tangible entity, defined in terms of psychological needs, not a mere statistical artefact, however elegant the procedure by which it is reached." Even Thomson describes the purpose of factor-analysis to be measuring "the factors of the mind," and objects to Spearman's system because it does not "give the *causes*" ([87], p. 64: his italics). For the rest most writers fall back on tacit identifications or equivalences. Thus Thurstone, Alexander, Holzinger, and many others treat the words 'factor' and 'ability' as synonymous [1]: but such an interpretation would rule out all application of factor-analysis to anything besides cognitive traits or tests [2]; others refer to them more broadly as 'elementary' or 'unitary traits of personality' (Kelley), or as 'the fundamental dimensions of the mind' (Guilford). Few, if any, explain why some factors are 'meaningful' and others merely 'statistical,' what makes one ability more 'fundamental' or more 'elementary' than the others, how to distinguish 'true psychological factors' from the rest, or the 'causal' from the merely 'descriptive.' Nevertheless, nearly all appear agreed that the factors sought by the factorist, however else they may be characterized, are at once real and causal. Thurstone, indeed, does not hesitate to

[1] E.g., [84], p. 54. Usually they speak of 'underlying abilities' (Holzinger, [106], p. 5), or of 'simple,' 'primary,' or 'fundamental abilities' (Thurstone, [122], pp. 1, 2, 4), again using the different adjectives as equivalent. Holzinger, however, incidentally makes an alternative identification, speaking of 'sorting mental tasks into distinct *categories or factors*' and of 'classifying traits according to correlation clusters' ([106], p. 4), which comes much nearer to my own view.

[2] In 1912, when I first applied methods of factor-analysis to assessments of emotional and temperamental tendencies, it was necessary to defend such an extension against a vigorous band of critics. But of late, particularly in America, there have been numerous attempts to factorize non-cognitive traits of personality. Hence it is surprising to find so many factorists still writing as though factor-analysis had been applied to nothing but tested abilities.

sum up the position by declaring that "the object of factor-analysis is to discover the mental faculties." [1]

Let us then proceed to inquire in what sense the factors disclosed by such analysis can claim both reality and causal efficacy.

(i) *The Reality of Factors.*—In spite of the apparent unanimity which the foregoing descriptions reveal, perhaps indeed because of it, more than one voice in recent discussions has warned the factorist against the temptation to 'reify and deify his factors.' With such a caution, I am in close sympathy. Yet, being addressed exclusively to the psychologist, it may seem to deny to his working concepts a validity allowed to those of other sciences. Thomson, for example, writes: "My contention is that no degree of real existence need attach to statistical entities like *g* . . . I do not think that *g* has any more real existence than a standard deviation; I do not believe that it is mental energy or any of the other things it has been guessed to be." He points out, rightly enough, that, if a mathematician claims to have proved "the existence of a *g* which is real," he merely means that *g* 'exists' as a solution to a certain algebraic equation, and that it is 'real' inasmuch as it does not contain the factor $\sqrt{-1}$; on the other hand, the "ordinary non-mathematical reader interprets this phraseology as meaning that . . . he, like everybody else, possesses a *g*, just as he possesses a liver." [2]

[1] [84], p. 53. Again at the very outset of his first chapter he proposes to postulate "abilities and their absence as primary causes of individual differences," and adds: "this implies that individuals will be described in terms of a limited number of faculties" (*loc. cit.*, pp. 45–6). Spearman, on the other hand, explicitly insists that factors must not be identified with 'faculties' or even with 'abilities' ([56], pp. 40, 222 *et al.*).

[2] [87], pp. 64, 84, 90. The mathematical theory of quantum physics would take a different criterion for existence. "The structural concept of existence is represented by an idempotent symbol" (Eddington, *The Philosophy of Physical Science*, p. 162; cf. *id.*, *The Relativity Theory of Protons and Electrons*, chap. xvi). In factorial work it can be shown that the selective operator representing a pure and independent factor is an idempotent symbol ([113], p. 160; cf. p. 264 below): and it is suggestive to recall that the equation of idempotency ($E^2 = E$) appears as a fundamental postulate at the outset of Boole's *Laws of Thought*. But, of course, the idempotency of a

Now, though I do not identify the general factor *g* with any form of energy, I should be ready to grant it quite as much 'real existence' as physical energy can justifiably claim; nor do I feel that such a concession would in any way conflict with Thomson's own explanation of *g* in terms of the sampling of a structureless aggregate of numerous neurones or their bonds. The solution of factorial equations is not the only thing with which the factorial psychologist is concerned: he may seek evidence for the 'reality' of his concepts in many other fields. Hence I should rather put the position in this way. The mathematical part of the factorial argument is for the most part deductive; and therefore, like all deductive reasoning, is admittedly unable to guarantee the reality of the results deduced. In such an argument the factors are notions postulated, not things directly observed or measured. As a mathematician the factorist merely asks: what are the fewest and the simplest postulates I can make in order to describe the phenomena I am observing? And from these postulates he tries to reconstruct the facts observed. But, when he has completed his reconstruction, he seeks to check and verify his inferences; and so turns again to the empirical world.

Deduction, whether mathematical or non-mathematical, can never prove the existence of the concepts with which it deals. Induction is therefore required at two stages: first, at the beginning, to suggest the initial postulates; and secondly, at the end, to see whether the empirical facts answer to the corollaries deduced.

This alternating procedure is not peculiar to factorial psychology. It is followed in nearly every quantitative science. In thermodynamics the physicist will first assume a perfect gas or a perfect engine, described by a few simplified postulates; he then deduces that the volume of a gas varies inversely with pressure or that with constant volume its specific heat is constant; finally he examines the results of experiments to see how far his deductions hold true of actual gases and actual apparatus. He would be surprised,

selective operator cannot guarantee any *concrete* 'existence' for the class 'selected.' However, it would take us too far afield to discuss what precise meaning (if any) can be attached to the vague term 'existence.'

I think, if we insisted that the temperature of the gas had no real existence, or warned him not to reify heat, and argued that the only real entities must be the moving molecules, which, as we had conceived them, were nevertheless quite unlike any actual molecules. Yet, from a severely philosophical standpoint, all these strictures would be fully warranted ; and we could safely add that the quantity called heat appeared simply as a solution to an algebraic equation, and not as a fund of energy that could be observed at first hand.

The student, reading Thomson's argument, takes the gist of it to be that Spearman's *g* cannot exist in the sense in which physical energy exists ; and his impression seems confirmed when Spearman himself admits that the identification of *g* with some such energy is only a tentative subtheory, and not essential to his main hypothesis. Both parties appear to overlook the fact that, if the truth be strictly told, even physical energy can claim no real existence. It is, indeed, a glaring instance of the faculty fallacy. We feel warmth, see light, hear sound, experience movement and resistance. To account for such effects the nineteenth-century physicist simply postulated an ' energy ' of heat, of light, of sound ; to explain the fact that a moving body does work he simply endowed the body with a ' capacity ' for such work ; worse still, when the energy does not manifest itself visibly to the senses, he invented yet another capacity that he called ' potential ' energy. Then, at a later stage of his argument, he declared, in terms like Spearman's, that there are not a number of such capacities, all specific and irreducibly distinct ; there is just a single general capacity, common to all ; the various forms of energy, which we originally classified according to the sensory experiences they apparently excite, are to be regarded as manifestations of one and the same indestructible energy, which is pictured as fundamentally kinetic.

No doubt, the physicist of to-day would be far less confident about the real existence of the energy thus postulated. He would probably declare that the question was irrelevant to his line of thought. The assumption of such a concept enables him to unify his science, to relate his various

measurements, and to construct rigorous deductive arguments in quantitative terms; all that interests him is to know that the conclusions of those arguments can be verified: never exactly—that he does not expect—but to a reasonably close approximation.

Similarly, in psychology, if the question of existence is to be pressed, I do not see why our factors should not be entitled to the same kind of existence—or, if you will, the same kind of non-existence—that is allowed to physical forms of energy. Thomson would deny that general intelligence exists as a real entity or as a genuine cause, but would accept it as a symbolic way of describing the relations between the ultimate neuronic elements; these alone he regards as the real entities and the true causes. But his arguments, so far as I can see, no more invalidate Spearman's main position than the kinetic theory of heat precludes us from talking of temperature as for all practical purposes 'real,' and treating it as a concrete existent whose amount (which is also a sample or average) can be conveniently measured.

Intelligence I regard, not indeed as designating a special form of energy, but rather as specifying certain individual differences in the structure of the central nervous system—differences whose concrete nature could be described in histological terms.[1] But in any case, whether it is a 'real'

[1] Spearman states that I identify *g*, not with intelligence, but with 'power of attention' ([56], p. 88), and "ascribe individual differences of ability to inequalities in 'power of attention'" (p. 341). He supports his statement by a paragraph taken from my first article; but this paragraph was meant to be interpreted in the light of the physiological hypothesis put forward in the succeeding pages. In a suggestive book just published, Maxwell Garnett also quotes the same passage, and seems at first sight to draw the same conclusion (*Knowledge and Character*, 1940, pp. 144–5). And from time to time the inference that "what we call intelligence is merely an effect of attention" has been cited by teachers and by educational journals with approval. Accordingly, since this notion has been so often attributed to me, I ought perhaps to explain a little more fully my view of the relation between the two.

Having shown (i) that a general factor enters into all cognitive processes and (ii) that this general factor appears to be largely, if not wholly, inherited or innate ([16], cf. [22]), I proposed to define intelligence as *innate, general cognitive efficiency* ([20], p. 95); and then endeavoured to suggest an innate and inheritable basis for this factor in the structure of the nervous system.

property in either of these senses or in some other remains a question that cannot be solved by factor-analysis as such.

The function of the nervous system is, in Sherrington's phrase, to facilitate efficient adaptation by its 'integrative action.' And its capacity for integration must depend on (i) the number of neurones, (ii) the number of branches of the neurones, and (iii) the systematic arrangement of those neurones and their branches. In the cortex of low-grade imbeciles, it is these three characteristics that are seen to be abnormal; and we may assume that minor variations of the same kind are to be found in all individuals. Thus, quite apart from experience, I conceive the neuronic structure of the brain to be laid down or organized differently in different persons, so that it 'facilitates efficient adaptation' in differing degrees. It is also to be noted that, in the same individual, the nervous tissue, like every other tissue (e.g. hair), has much the same general structural character throughout. Consequently, this general structural character of the nervous tissue will operate as a 'general factor.' With these facts in mind I went on to argue that "that mind will, therefore, be most intelligent in which the tendency to *systematization* in the portions of the brain as yet unsystematized is greatest, i.e. where the inborn capacity for *complex organization* in the nervous tissue as yet unorganized is most rich: such a capacity can be pictured as dependent on the general architecture of the central nervous system, and, being structural, can, like all structures, be conceived as inheritable" ([17], 1909, p. 23; cf. [16], p. 169). Thus, like Thomson ([132], p. 54)], I consider that "persons differ widely in the 'richness' of their brains," but I lay more stress on the *systematization* of 'bonds' (as he and Thorndike term them) than upon the mere *number* of 'bonds.'

Now, if we accept McDougall's view of the attention-process (*Mind*, XII, pp. 302 f.), we may regard this integrative, co-ordinating, systematizing activity of the nervous system as the essential 'physiological factor' underlying attention in its cognitive aspect. Along these lines, it seemed to me possible to reconcile the views expressed by Wundt and Binet, who held that intelligence or intellectual power consisted essentially in attention, with the views of other writers, who had variously described intelligence as power of adaptation, capacity for learning, capacity for noetic synthesis, or the like. This suggested reconciliation, however, was merely a corollary to my conception of intelligence, not the primary definition of it. Indeed, instead of maintaining that 'intelligence is merely an effect of attention,' I would rather say that attention (so far as it is a cognitive and not a conative phenomenon) is an effect or symptom of what we call intelligence.

Spearman, it may be noted, would agree to the identification of *g* with attention, if attention meant, not 'intensity of conation,' but 'mental energy': McDougall, on the other hand, would prefer to identify 'mental energy' with 'intensity of conation' and attention with the *direction* of energy or conation. Garnett's final way of putting the matter I should readily accept: "*g* measures my power to concentrate attention when I wish to do so; *w* is the measure of the influence of my purposes on my use of this power" (*loc. cit.*, p. 150).

Just as we cannot deduce the essential character of the elementary processes in the retina from a mere analysis of colour equations, so we cannot determine the nature of intelligence without supplementary evidence from anatomy, physiology, and genetics, to be accumulated by research that is only just beginning. Meanwhile, if (as many factorists affirm) it is unfair, in our present state of ignorance, to import physiological considerations into the picture, and if we are consequently to take our factors as describing, not hypothetical characteristics of the individual nervous system, but only observable characteristics of the individual mind, then certainly we had better refrain altogether from referring to such factors as concrete entities: to speak of 'factors in the mind' as if they existed in the same way as, but in addition to, the physical organs and tissues of the body and their properties, is assuredly both indefensible and misleading.

Strictly, therefore, the scientist can never really measure mental abilities as entities in themselves; for there is no ground for believing that such abilities can have any real existence apart from the behaviour by which they are displayed or the organism that displays them. What we call an ability is simply a convenient name for designating a set of potential reactions on the part of the individual tested.

(ii) *The Causal Efficacy of Factors.*—If our mental factors as such can claim no necessary concrete existence, still less can we endow them with effective causal powers. Explanation in terms of causal agencies is a legacy of nineteenth-century science—a mode of approach which should now be considered as out-of-date in mental science as it is in physical science. In psychology concepts of this kind appear little more than relics of the old-fashioned faculties, which were invented to perform the same function in psychology as were performed by forces in physics, and have since got ingrained in our everyday habits of speech. Nearly all psychologists are nowadays agreed in repudiating mental faculties, or at any rate the name; but their reason for this rejection seems rather that there is little or no objective evidence for any set of faculties that has hitherto been

proposed, and not that the whole idea of faculties springs from a crude and obsolescent notion of scientific explanation. Factor-analysis, I believe, owes much of its present confusion to the fact that most psychologists still tacitly assume that, if faculties do not exist, mental factors must be invoked to fulfil their explanatory functions.

The causal phraseology, which almost every factorist continues to employ, implies a naïve and popular view of the nature of mind. The minds of the persons tested are conceived as individual substances; and the 'abilities' inferred are then pictured as causal properties inherent in those substances. The underlying motive is not hard to discern. If we wish to predict a future condition from the present, we seem compelled to assume some principle of constancy or conservation, an entity permanent enough at least to last until the date to which our predictions refer; and by the unsophisticated thinker such a principle is nearly always visualized concretely as an enduring substance or as part of such a substance. Yet even in explaining physical phenomena, the hypothesis of individual substances possessing causal properties or attributes would scarcely pass unchallenged by the metaphysician: and in psychology, where we are no longer dealing with separable bodies of matter, but, for all we know, merely with certain aspects of the working of the nervous system and its adjuncts, such a way of speaking is even less permissible than in other fields of science. It manifestly begs a host of unsolved metaphysical issues.

Philosophers and psychologists alike have continued to suppose that science must aim primarily at discovering causes, although as a matter of fact the word 'cause' vanished long ago from the vocabulary of the more advanced of the sciences. The philosopher regards causality as a weakness of physical science; the psychologist, as its special strength. Yet nowadays in the physicist's account of gravitation there is nothing that can be called either a cause or an effect: there is only a quantitative law embodied in a mathematical formula. Already by the end of the nineteenth century,[1] causal interpretations in physics were

[1] The extrusion of causality from physical science was virtually accomplished by writers like Kirchhoff and Mach abroad and Karl Pearson in this country. With the general recognition that force and components of force are mere mathematical fictions, causality disappeared. It lingered longest in physical dynamics, and still persists in the psycho-dynamics of

beginning to give place to differential equations; thermodynamics was frankly statistical; and in quantum physics, individual substances as such and causal properties as such find no mention whatever. The modern physicist would emphatically disclaim any knowledge of underlying entities or of their possible powers: he finds only an intricate and abstract structure, a pattern of observable relations between observable phenomena—the phenomena themselves being either mere patterns of relations, or else patternlike sequences of events within a conscious continuum, and so neither substantial nor causal, and certainly outside the purview of the physicist as such.

This attitude seems now to be shared by nearly all contemporary writers, even those whose metaphysical views are as sharply opposed as the views of Eddington and Russell. "Physics to-day," says Eddington, "represents experience as the result of statistical laws without any reference to the principle of causality. We seek a knowledge, neither of actors nor of actions, but of a structure or pattern contained in the actions." [1] Similarly, Bertrand Russell insists that many philosophical difficulties "might have been avoided if the importance of structure and the difficulty of getting behind structure had been realized. . . . The essence of individuality always eludes words and baffles description, and is for that very reason irrelevant to science." [2]

Why, then, has the so-called law of causation persisted so long in the writings of psychologists, and even been magnified, in the doctrines of the behaviourist and the psychoanalyst, into an axiom of supreme importance? Possibly

those behaviouristic and psycho-analytic writers who adhere to materialistic theories already out-of-date in the fundamental material sciences. The opponents of causality usually ran to the opposite extreme, and insisted that all science is merely descriptive and that in the narrowest sense. There seems a *via media*. Without accepting the extreme causal view, on the one hand, or the extreme descriptive view, on the other, we may adopt the position that, although causal links are not discoverable, *implications of probability* are, and these will enable us to predict as well as to describe without assuming any external causal compulsion.

[1] *New Pathways in Science* (1935), p. 256.

[2] *Introduction to Mathematical Philosophy* (1919), p. 61.

because psychologists are unfamiliar with any other form of law. Until recently, the idea of a functional law as distinct from a causal law seems to have been quite foreign to them; the nature of correlation, for example, has nearly always been explained to the novice in terms of causes rather than of dependence. Yet, after all, the most that science requires for its probable inferences, whether theoretical or practical, are not so much laws of cause and effect as *laws of ground and consequence*. What we need to know is not the cause of an occurrence, but the reason for a conclusion. When we say 'A *causes* B,' all that we mean is 'A's existence at the time of determination *implies* B's existence at the time to which our predictions refer, provided other conditions do not appreciably interfere.' Our questions therefore must be, not what conditions compel such and such things necessarily to happen, but what conditions enable us to infer that they will probably happen. For this reason, as it seems to me, the notion of logical *dependence* lies at the very root of factor-analysis in all its forms and all its applications; and for this reason the more special problems of linear dependence and statistical dependence are always coming to the fore.

The Postulates of Inductive Inference.—When this view is put to the empirical psychologist, I find he is usually ready to admit it for what he calls the deductive sciences; but he still insists that laws of causation are necessary for psychological science, because it is essentially inductive. Without the fundamental premiss of causality, he supposes, there can be no valid induction. The modern logician, however, would at once point out that reasoning by causes is popular reasoning, not rigorous reasoning, and that neither the law of universal causation, nor yet (what has frequently been identified with it) the axiom of the uniformity of Nature, is necessary or even sufficient to guarantee the validity of inductive proof.[1]

[1] The only logician quoted by factorists seems to be Mill. His law of causation, his canon of induction, together with the principle of parsimony or simplicity (wrongly attributed to Occam), are adduced again and again. Until recently, indeed, not only empiricist logicians of the school of Mill, but even rationalist logicians like Joseph, would have agreed that a law of

Nevertheless, some *a priori* postulate is avowedly requisite. And the appropriate postulate or postulates may in their turn throw light upon the fundamental nature of the subject-matter about which we desire to reason. So far as physical phenomena are concerned, most logicians would probably prefer to invoke, in the place of the traditional law of causation, something more akin to the two postulates proposed by Keynes, which he terms the 'principle of atomic uniformity' and the 'principle of limited independent variety.' [1]

Strangely enough, the very philosophers who accept these two principles for the physical sciences warn us that they "do not apply to inductive generalizations about *mental* phenomena; so that with our present knowledge we have no good reason to attach great weight to the conclusions of inductive argument on these subjects" [2]—a shattering verdict for those who seek to reduce mental phenomena to something like an empirical science. Yet I venture to affirm that these two postulates, or something very like them, are precisely what are needed to justify the inductive inferences of the factorist. And if that view is right, we could, it seems to me, contend that factor-analysis may in this way throw an illuminating beam on the ultimate structure of mind, or at any rate on the best work-

universal causation was needed to help us out of the apparent fallacies that all inductive arguments otherwise seem to involve, and further that it could do so successfully. To cite the arguments brought by more recent logicians against such a doctrine is hardly necessary here. Of the many refutations perhaps the most readable and forcible is that set out by Russell in *Mysticism and Logic*, chap. ix.

[1] J. M. Keynes, *A Treatise on Probability*, pp. 256 *et seq.* A modified version of them is elaborated by C. D. Broad, whose exposition will probably be found a little clearer and more suggestive to the factorial psychologist. ('Principles of Problematic Induction,' *Proc. Arist. Soc.*, N.S., XXVIII, pp. 1–47, and 'The Relation between Induction and Probability,' *Mind*, XXVII, pp. 389 *et seq.*). A still more recent statement of these views in *Le problème logique de l'induction*, by Jean Nicod (ably summarized by R. B. Braithwaite in *Mind*, XXXIV, pp. 483 *et seq.*), contains several important revisions. But even if he hesitates over trying to master the current views of inductive logic, every factorial worker should at least be familiar with Keynes' *Treatise*.

[2] Broad, *loc. cit.*, p. 46.

ing conception that can be adopted in regard to that structure. What is more, the particular conception thus suggested would appear to be, not merely consistent with, but actually confirmed by, recent views on the general character of mental and neural phenomena themselves.

The first assumption is, as Keynes points out, a generalization of the principle known to mathematicians as the 'superposition of small effects.' It maintains that the universe and its processes may be treated as consisting of quasi-atomic elements, so that "a change of total state may be considered to be compounded of a large number of smaller separate and independent changes."[1] A mental change, for example, might be regarded as the resultant of numerous all-or-none discharges of certain nerve-cells. The relevance of Keynes' postulate to current statistical reasoning is plain. In factor-analysis it supports the demand for an ultimate analysis into independent or 'orthogonal' factors; and it would seem to form the implicit basis on which Thomson's arguments for a 'sampling view' of factorial problems must really rest. It is an assumption that colours the whole outlook of psychologists who belong to the analytic and determinist school. Behaviourists like Watson, associationists like Herbart, Titchener, and Thorndike, determinists like Freud, have explicitly invoked some such atomistic principle—usually in what would now be regarded as a crude and untenable shape. The factorist has similarly been accused by the Gestalt and Intuitionist schools of clinging to an atomistic view of mental process that "inevitably disrupts the personality into separate bits."[2] However, if, in Keynes' formulation of his principle, we substitute the word 'distinguishable' for 'separate,' such criticisms lose much of their force; in any case, neither analysis nor inductive inference as ordinarily stated seem able to dispense with some such postulate.

[1] Keynes, *loc. cit.*, p. 249. In his recent Tarner lectures Eddington maintains that the mode of analysis in physics rests on what would seem to be a very similar principle—viz. 'the atomic concept or the concept of identical structural units.' He regards this as a necessity of the logical framework of our scientific thinking (*The Philosophy of Physical Science*, 1939, p. 122).

[2] Cf. Vernon, *Character and Personality*, IV, pp. 1–10, and Spearman's reply, *ibid*, pp. 11–16.

Nor does it stand alone. We need, as we have seen, a second principle. Briefly this may be expressed as follows: "the almost innumerable observable characteristics of any object may be treated as deducible from a smaller and finite number of independent generating factors": so that "the objects in the field, over which our generalizations extend, do not have an unlimited number of independent qualities: i.e. their qualities, however numerous, cohere together in groups."[1] Thus stated the principle may be regarded as a refinement of the traditional doctrine of Natural Kinds, Real Species, or Types.[2] It is complementary to the preceding; and warns us that the postulate of atomic uniformity must be qualified by observing that the atomic elements are not absolutely isolated, but join to form systematic sets or wholes, each more or less continuous within itself, and not always completely discontinuous from others. Neural cells are integrated into patterns or systems: mental traits seem to hang together in clusters; persons can be classed together to form what used to be called types; sensations coalesce to form Gestalten, or rather (for here it is fallacious to put the part before the whole) sensations are simply distinguishable aspects in complex and continuous patterns. Together, the two principles allow us to treat the phenomena to be factorized as consisting, structurally, of systems of correlations superimposed upon a background of non-correlation or 'chance' that is essentially structureless. The upshot is that the apparently unlimited number of determinates that are actually observable may be regarded as arising from the relations between a comparatively limited number of determinables.

A 'Natural Kind,' says Broad, 'is a region containing a

[1] Keynes, *loc. cit.*, p. 25. I take it that the term 'generator' has a mathematical meaning and was perhaps suggested by Laplace's doctrine of generating functions. In any case, as we shall see in a moment, the notion of a 'generating factor' is not so essential to the principle as the notion of a 'coherent group.'

[2] See Mill, Bk. I, chap. vii, pp. 134–6. Bk. III, chap. xxii, p. 107. "Among the uniformities of co-existence which exist in nature may be numbered all the properties of Kind. . . . A portion of them . . . are independent of causation"; but "it is impossible in any case to be certain that they are." Cf. also *ibid.*, pp. 260 *et seq.*, and p. 69 above.

blob. . . . What we find is not a regular distribution of all the actual states among all possible states, but a bunching together of instances in the neighbourhood of certain sorts of states.'[1] Factor-analysis might well be described as a method of locating such ' blobs.'

Thomson, as we have just seen, and with him many of the more cautious writers (Thorndike, for example, and M. S. Bartlett), have been led to assume that the mind must be a relatively structureless aggregate of similar elements independently acting and unlimited in number.[2]

[1] *Mind*, *loc. cit.*, p. 25. It may be noted that Keynes' first principle was intended to rule out the possibility that ' natural law might be organic and not atomic ' (*loc. cit.*, p. 249) ; but the second seems to allow us to introduce ' organic ' relationships. Hence, I should prefer to say that all we need is to assume that natural law, even if organic, may be treated as atomic by way of a first approximation. However, what is still more interesting is the fact that Broad seems to demonstrate successfully that, once we have used the first principle (in the form of elementary generators) to reach the second, we can drop it out, and take the second (in the form of mutually exclusive coherent sets of observable characteristics) as alone being fundamental. " The hypothetical generating factors can now be regarded as no more than convenient parameters : they *may* exist, but it is not necessary to suppose that they *do*. . . . It must therefore be possible to eliminate them, and to state the case wholly in terms of observable characteristics and their relations " (*loc. cit.*, p. 39). This, as we have seen, is precisely what the matrix formulation of the problem enables us to do in detail for any particular instance.

[2] The empirical reasons for this conception are apparently that the nervous system is built up out of innumerable cellular units and that behaviour is built up out of innumerable reflex bonds. If some kind of structural organization appears later within the mind, that, it is argued, is " probably because education and vocation have imposed a structure on the mind which was absent in the young." In considering this argument two points seem pertinent. First, even within the cortex the cellular elements are not innately identical in form or function, nor have they equal and unlimited connexions with all parts of the nervous system (including its sensory and muscular appendages). Secondly, from the very outset the young organism reacts as a whole to its environment as a whole : the specialized responses, though in part innately determined, mature gradually within this integral mode of behaviour ; and the traditional ' reflex action ' of the textbook is not the unit out of which behaviour is built up, but a comparatively late feature in maturation (cf., for example, Coghill, ' The Structural Basis of the Integration of Behaviour,' *Proc. Nat. Ac. Sci.*, XVI, p. 637 f.). This conception is in keeping with the correlational results obtained on testing children of different ages : as I have pointed out elsewhere, " the relative influence of the general factor is greater in earlier years as contrasted with later." " Group-

Yet, taken rigidly, this principle, if it stood alone, would seem destructive of all factorial work, as commonly conceived. And Thomson himself has recently introduced a second qualifying assumption. He now assumes a mind that is not entirely structureless, but divided into ' regions ' or ' sub-pools ' within the total ' pool.' This further assumption, as it seems to me, is really a special application of our second postulate, namely (as it manifests itself in psychology) that both mental traits and individual minds exhibit not a regular or homogeneous distribution, but a limited variation with a limited independence, a confluence of individual traits into ' group-factors ' and of individual persons into ' types '—in short, a mottled distribution into what Broad calls ' naturally cohering sets.' [1]

We have noted at an earlier point (p. 93) how the elaboration of the method of matrix analysis by the modern psychologist, on the one side, and the modern physicist, on the other, and their simultaneous and successful employment within their respective spheres, suggests that as regards ultimate logical structure both the mental world and the physical world must be fundamentally akin. Here, as it seems to me, we have the ground of that kinship made explicit. Whether this common structure in its turn has been imposed on both worlds *a priori*, like a set of Kantian categories, by the logic and the laws and forms of thought which every human analyst is forced to use, is an epistemological problem into which we need not enter. The main

factors are far more conspicuous at the stage of vocational guidance than at that of educational guidance " ([41], p. 266; cf. [35], pp. 63 f.).

As regards the statistical and more general arguments for the conception I have just criticized (and these are the arguments on which Thomson and Bartlett chiefly rely) the difference between us would seem to be mainly one of emphasis (cf. [137], p. 88 f., and *Brit. J. Educ. Psych.*, IX, pp. 191 f.).

[1] For the philosopher, I suppose, the problem is the ancient puzzle of reconciling what Plato called the Many and the One. Certainly it is not peculiar to psychology. " One of the most remarkable achievements of current quantum theory is the way it has surmounted the difficulty of giving to the parts of the universe a kind of self-sufficiency, which does not cut them off from interaction with the rest." (Eddington, *loc. cit.*, p. 127.) This achievement, we are told, has been rendered possible by a mathematical technique, which, as we have seen (p. 165), is closely similar to that which has been advocated here.

conclusion seems clear. Rigorously speaking, factors cannot be regarded as substances, or as parts of a substance, or even as causal attributes inhering in a substance. They are not separate ' organs ' or isolated ' properties ' of the mind ; they are not ' primary abilities,' ' unitary traits,' ' mental powers or energies.' They are principles of classification described by selective operators. *The operand on which these operators operate is not ' the mind,' but the sum total of the relations between minds and their environment.* The relational structure of this operand the factorist must presume to be knowable, but its causal or substantial nature he must treat as unknown or at any rate irrelevant.[1]

Three Levels of Interpretation.—Without launching into the more elusive problems of epistemology, we may, I think, clarify and harmonize the conflicting views put forward on the reality and causal efficacy of mental factors, if we recall the three distinct aims of reasoned analysis—prediction, explanation, and description. These broadly correspond to the three ways in which man has progressively approached the study of the world around him—the practical, and then the philosophical, and finally the scientific. In sciences that do not deal with man himself the non-scientific issues are less prone to obtrude. The botanist, as he classifies his field specimens, does not worry whether herb robert would develop better if transferred to garden soil, or speculate whether the beauty of each fading rose is transcendentally immortal. But the psychologist, particularly if he is also a factorist, is continually running into problems that are not scientific at all, but sometimes practical and sometimes metaphysical. He cannot avoid them. For him, therefore, it is of supreme importance to keep the different planes of thinking separate.

The practical thinker is interested chiefly in forecasting the effects of alternative courses of action : the medical psychologist at the clinic, for example, wants to know what to do with his patient in the near future, and what will be the ultimate outcome ; like most practical workers, he feels that,

[1] I add irrelevant, because it is conceivable that the intrinsic nature and causative operations of the mind might be revealed by introspection or tentatively deduced by metaphysical speculation.

unless some degree of causal determination is presupposed, his anticipations can have no basis. The philosophical thinker searches rather for explanatory realities behind the immediate facts. In dealing with mental phenomena, we are all of us tempted to philosophize. But when the practical man begins to do so, particularly if (as with the medical worker) his practical training has left him unaware of the pitfalls that await the amateur philosopher, he is prone to carry over his practical concepts into the realm of metaphysics, regardless of the fallacies that such a procedure entails. An austerely scientific attitude comes last of all. The scientist is, or should be, the most cautious of thinkers. His concern is solely with the systematic description of his own restricted province. The co-ordination of his facts will necessitate inference : but his inferences must rest as much as possible on the facts themselves and as little as possible upon assumptions. His method, therefore, will be, not to guess at supposed realities to satisfy some practical or philosophical need, but ruthlessly at every step to eliminate whatever is neither proved nor requisite for proof. Let us remember, then, that a type of explanation that may be temporarily helpful on the lowly level of applied or practical psychology may be both a snare and a delusion on the higher level of metaphysical speculation, and can only be accepted on the intermediate level of pure or empirical psychology if it has stood the sternest scrutiny.

(*a*) *On the Level of Applied Psychology.*—In practical psychology, as distinct from theoretical, there is some excuse for causal language. In the field of educational, vocational, and clinical work, the logical grounds that the psychologist seeks are grounds for inferences specifically in regard to the *future*. Now a cause, at any rate in the popular mind, always precedes its effect in time. If, therefore, we postulate a law of causation, and search for these anterior causes, we may be able to deduce what will occur in the future from conditions that are ascertainable now. Instead, then, of a symmetrical or reciprocal dependence, such as could be expressed by an abstract function, the practical man is interested only in the asymmetrical or one-sided dependence of the future on the present or the past ; and this one-sided

temporal dependence is what he seeks to formulate when he talks of causal factors. His motives are clear. He is thinking primarily about the outer world ; and it would perplex him if he were obliged, as the scientist and philosopher are obliged, to keep thinking also about the validity of his thinking. Consequently, for the logical concepts of ' ground and consequence,' of reason and conclusion, he substitutes the more concrete notions of ' cause and effect.'

Nevertheless, the language of causation, particularly in psychology, brings with it two misleading implications. First of all, causal knowledge is supposed to be certain knowledge : *vere scire est per causas scire.* But the very fact that our data, by hypothesis, supply knowledge of antecedent conditions only—often of remotely antecedent conditions at that—implies that our so-called ' causes,' as known, can cover only the incomplete ' ground,' not the sum total of necessary and sufficient conditions, as is erroneously assumed ; and the shortcoming proves particularly baffling in the study of mental behaviour, since here future possibilities—for example, what a man desires or intends to do—are frequently required to explain his actions, and may explain them far more clearly than present or past conditions by themselves. It follows that all inferences from causes, particularly where the ' ground ' is highly complex, and most of all where it is partly mental, are bound to be, not certain, as is commonly imagined, but merely probable. Indeed, when a psychological determinist—an upholder of psycho-analysis or behaviourism, for example—trusts dogmatically to the postulate of complete causal determination, the conclusions that he proceeds to deduce are likely to have a low rather than a high probability. And much the same danger besets the factorist : no predictions are so confidently offered by the educational or vocational psychologist as those which are based on some such factor as *g* or general intelligence. Yet again and again his forecasts are falsified. A broader clinical approach, though yielding more tentative and provisional predictions, would have saved him from many pronouncements, that may seem to possess precision of a narrow mathematical sort, and yet ultimately serve only to discredit his methods in the eyes

of the practical teacher, who relies more on prolonged first-hand experience than on precise academic deductions.[1]

Secondly, the use of the term cause implies, not merely logical necessity, but physical necessity, i.e. that the action will be the result of a quasi-physical force, ' forced ' in the sense of compulsory. This leads to a still cruder type of determinism—a type which is far more undesirable, and certainly more illegitimate, in psychology than elsewhere, since it persuades us to regard human actions, not merely as intelligibly determined, but as mechanically determined. For psychological guidance or treatment it is no doubt convenient to think of mental growth and intellectual progress as dependent, at any rate in part, upon certain constants that remain stable throughout the individual's life ; and, in view of their correlations with similar features in other members of that individual's family, it is even permissible to speak of these constants as (in a rather loose sense) inherited or innate. The search for factors thus becomes, to a great extent, an attempt to discover inborn potentialities, such as will permanently aid or limit the individual's behaviour later on : and in the results of our tests we therefore try to sift and separate different hereditary capacities both from each other and from the effects of

[1] It is the combination of the ' two-factor theory ' of Spearman with the ' biological determinism ' of Watson that has led, in the practical applications of their joint followers, to those rash inferences about individual children which Mr. J. C. Hill so strongly deprecates (see p. 54 above). Spearman himself, I believe, has never drawn the sweeping inferences of which Hill complains. Indeed, it was precisely to escape such criticisms that he introduced the designation *g* in place of ' general intelligence.' But others have certainly been tempted to forget that a child's ' factor-measurement for *g*,' as directly computed from a set of tests, may undoubtedly have been affected (as Hill suggests) almost as much by " the poverty of his home environment " and by " the influences of the first five years of life " as by innate or hereditary conditions. At the same time I see no reason to run to the other extreme, and declare with Hill that ' *no* conclusions about the distribution of innate ability can be drawn from the figures ' (*loc. cit. sup.*, p. 271). His own inference is quite as illogical as the inferences he attacks. He might just as well argue that, because the influence of the sun can never be eliminated when we are studying the orbit of the moon, therefore no conclusions about the earth's influence can be drawn from the figures observed. It is the object of mathematical analysis to separate the two influences so far as possible and to state the relative weight to be attached to each.

experience or training that we assume to be superimposed upon them.

This artificial subdivision of the mind into independent components greatly simplifies the task of prediction; yet, unless we continually remember that our predictions can never yield more than probabilities and first approximations, we shall make the fatal mistake of treating the individual's potentialities as far more rigidly fixed than is actually the case. In educational work the doctrine of an innate general factor of intelligence has, I believe, been on the whole helpful rather than harmful. But, as applied to adults rather than to children, it involves great dangers. With both, the notion of innately limited abilities needs to be employed with far more caution than hitherto; and the identification of temperamental factors with such concrete physiological influences as physical types or endocrine secretions, over which the individual has no control, though full of suggestive and neglected possibilities, has led both factorial and non-factorial writers to rash speculations, and has often proved a needless obstacle to proper clinical treatment.

Nevertheless, I would not be so pedantic as to banish causal terms altogether from the psychologist's vocabulary. The physicist, who would never mention causes in a book on quantum physics, would not scruple to use the word in chatting over a practical problem with an engineer. Similarly, in practical or applied psychology, where, for example, we are canvassing the history and the handling of subnormal cases, the language of causation is not only convenient, it is almost unavoidable, if we are to remain comprehensible.

As a practical psychologist, then, though not as a theoretical psychologist, I should consider myself licensed to talk in terms of causal factors—abilities, temperamental tendencies, acquired habits, and the like, just as a modern astronomer is still free to talk of sunrise and sunset. And I should certainly prefer to express these causes in genetic, physiological, or biochemical terms, were only our knowledge sufficiently advanced. Since it is not, I have to give my causes mental rather than material labels, and speak of

verbal ability rather than of a verbal centre, of instincts, habits, and even unconscious wishes rather than of this or that neural mechanism, of general emotionality rather than of neural energy or power. But I still try to keep these concepts parallel, so far as may be, with what is presumably the organization of the underlying physical basis, and consistent, so far as possible, with existing knowledge of the working of the nervous system.

(*b*) *On the Philosophical Level.*—As a philosopher, too, I should again be willing to open up the possibility of mental causation, and even to inquire whether there may not be after all some justification for conceiving the individual or his mind as an independent substance. But the substances and the causes that I should then envisage would not be the homely substances or causes that I ingenuously refer to as a clinical psychologist or as a lecturer to teachers or medical students.

This is not the place to embark on a full metaphysical disquisition. A comment or two must suffice. In brief, the philosophical theory that I should offer would not be very far removed from the assumptions to which I seem directly driven by the immediate exigencies of factorial work. Roughly, it might be described as a modernization of the old Platonic doctrine of εἴδη or 'ideas.'[1] Its main principle would be that reality is best described in terms of 'forms,' 'structures,' or *Gestalten*—things analogous to the cognitive wholes that we perceive in our own personal consciousness, but also possessing something of the causal efficacy that we seem to find on the conative side of our experience. And I should argue that, whether we are

[1] Elsewhere I have drawn attention to the remarkable way in which some of the most characteristic implications of factor-analysis, Gestalt psychology, and quantum physics seem to have been anticipated in certain passages in Plato and Aristotle (cf. the *Note on Faculty Psychology*, written for the Consultative Committee of the Board of Education and published in the Spens *Report on Secondary Education*, 1939, pp. 429 *et seq.*). The analogies are perhaps important as reminding educationists and others that modern psychological opinion is not, as is so often supposed by the layman, tending at the moment in a materialistic or fatalistic direction, but is apparently compatible with the highest kind of idealism that has inspired social and educational efforts in the past.

studying the material world as physicists or the world of experience as psychologists, the only articulate or communicable knowledge that we can attain is a knowledge of structure. Let me add that, if this position be accepted, the relation of matter and mind would lose much of its mystery, for we should no longer be concerned with the interactions of disparate substances but with the correspondence of abstract structures.[1]

At the same time, to save misunderstanding, let me add that I do not regard physical phenomena and psychological phenomena as *entirely* on the same footing. Physical processes must be described, and can only be described, in terms of relations and systems of relations. But this is " because " (to quote a phrase used by Johnson in a somewhat similar context) " the ultimate constituents of matter—if there are ultimate constituents—have, so to speak, no insides." [2] The ultimate constituents of consciousness, on the other hand, are ' insides ' that we know at first hand, even if that knowledge is not as such directly communicable. General psychology, therefore, which includes introspective psychology, has to take this further feature into account. Factorial psychology, being concerned with the psychology of others, whose ' insides ' are not directly known to us, must adopt a more contracted standpoint. But these reservations are scarcely relevant to my main argument, and are inserted only to forestall a possible criticism that their omission might have provoked.

(*c*) *On the Level of Empirical Science.*—As a philosopher, therefore, I should have no desire to shirk the deeper problems of causality or the conception of the mind as an ultimate entity or substance. But as a mere scientist I have no right to express an opinion on such issues, much

[1] This would simplify the problem of perception : for, although nothing can *resemble* an idea except another idea, a material environment can *correspond* to a perceptual continuum ; like a map and a landscape, both may have the same relational structure, and that is the only kind of ' resemblance ' we require. Again, it would partly elucidate the relation of brain to consciousness : for the changing field of consciousness may correspond with the changing field of cerebral tensions. Indeed, many of the problems of Gestalt itself and of unconscious cerebration as well as of cognitive activity could be very simply interpreted if we regarded the brain as a kind of organic machine for continually solving matrix equations.

[2] Johnson, W. E., *Logic*, Pt. III, p. xxiv.

less to commit myself, openly or tacitly, to any particular solution. Pure psychology is concerned solely with inquiries on the intermediate level, below the loftier plane of the metaphysician, who alone is authorized to talk of realities and to include or exclude souls and their transcendental powers, but above the humbler plane of applied psychology, where the everyday parlance of individual substances and persons, of cause and effect, of abilities and will, again becomes permissible. In pure psychology, as in pure physics, the word 'causes' ought in every rigorous argument to be carefully eschewed, and be replaced by some non-committal description such as that of 'functional relations.' The psychologist, as it seems to me, must accept the warning which Eddington has addressed to the physicist, namely, that "so-called causal events are to be thought of merely as conspicuous foci from which the links radiate," in short, as the centres of naturally cohering sets; the 'links' are not "lines of force or power," but simply "determinate laws or mathematical equations connecting the events"; and it is "with these links or relations only, and not with the relata, that the humble scientist is concerned."

Relativity in Psychology.—The very terms 'co-relation' and 'co-variance' ought to have suggested at the outset that we were about to analyse relations between qualities and variations in qualities (that is, differences between qualities), and were no longer dealing with qualities themselves. No one has ever thought of the variance or the standard deviation of a group as measuring a mental ability or power: and hence, I imagine, no one would regard the covariance of a group as measuring a mental ability or power. Now, by its formula, a coefficient of correlation is nothing but a covariance, arbitrarily standardized. Unfortunately, however, in many psychological textbooks a coefficient of correlation is described as a measurement of a tendency. As a result, the reader insensibly substitutes the idea of a single measurable thing, the 'tendency,' for the idea of a 'constant' (and a very inconstant constant) summarizing a system of measurements; and the student is apt to believe that a correlation coefficient describes some-

thing more real than the covariance, whereas the reverse is the case.

Thus, although most psychologists treat factor-analysis as essentially concerned with the analysis of correlations, I myself believe that it is better regarded as a mode of analysing variance. Variance is admittedly a statement of variation—that is, of differences, and of differences only. And this mode of treatment, imposed on us at almost every turn, brings out an important peculiarity of psychological measurements as contrasted with physical measurements, or at any rate with the common conception of physical measurements. When we measure children's heights and weights, our figures have the form of absolute measurements taken from a zero point that is extraneously determined. In psychology a few of our measurements, it is true, also appear in an absolute form, e.g. the mental age or the speed of reaction. But, whatever be the shape of the initial measurements, in calculating a correlation our very first step is to convert those measurements into deviations from an intrinsically determined zero, namely, the arithmetic mean.

The same is true of the factor-measurements and the factor-saturations derived from those correlations. The final factor-measurements as ordinarily computed are expressed in standard measure : that is, they too are deviations about their own mean. If, as I believe, the factor-measurements for traits (determined by correlating persons) are equivalent to the standardized factor-saturations for traits (determined by correlating persons), then even the saturation coefficients should in general be expressed as positive and negative deviations. Thus, a factor-measurement can never specify the absolute quantity of ability that a man is supposed to possess, as we might specify the pints of blood in his system or the ounces of brain-tissue in his skull : it can only express the *relations* between different persons' performances. And similarly the factor-saturations can only express the way individual *differences* in test-performances or measurable behaviour depend on individual *differences* of a more general kind.

Sometimes, it is true, correlations are said to measure

resemblances. But we now see that it would be equally true, and far less misleading, to say they indicate *lack* of resemblance, or difference. It would be truer still, and much more instructive, to think of our tables of measurements, correlations, factor-saturations and the like as comprising a series of mutually equivalent matrices, each capable of being transformed into the other, and to note that each of those matrices, even if ultimately reduced to a single row or vector, can still enumerate only relations between qualities and not the amounts of those qualities by themselves : just as the co-ordinates of space and time can only state the position of a star in regard to some other object or observer, and never its absolute position in the universe.

If, however, we adopt the commoner view—that factors are ' powers ' or ' abilities ' or ' mental energies,' which can aid performance but never hinder it, and that the factor-saturation measures the causal influence of the ability on the test, and that factor-measurements state the amount of ability that a person has at his disposal—then we shall be forced, with Spearman, Thurstone, and Alexander, to accept a far narrower view of the possibilities of factorial work. With them we shall be compelled to reject the whole notion of negative saturations : for with them we shall argue that a factor, being an ' ability ' or ' organ,' can only be present or absent, never a minus quantity ; hence it can have only positive or zero saturations (precisely zero, not approximately zero).[1] Wherever we find any negative values, we shall start ' rotating our axes ' until we are left with positive saturations only. This will mean, paradoxically enough, that, having begun with a set of correlated tests or traits, and having reduced them to terms of independent

[1] See below, chap. xii. I have defended the acceptance of negative saturations elsewhere ([137], p. 90) and need not repeat the arguments here. Once again, a similar change of standpoint is discernible in physics, namely, from concrete explanatory principles, which permitted only positive effects, to wider and more abstract principles that would permit negative relations as well as positive. " The concept of substance introduces . . . a limited form of analysis . . . in which the systems are restricted to those which furnish a set of *positive* parts. . . . Though there may be such cases in physics . . ., we now look on it as an incidental restriction in a particular application " (Eddington, *loc. cit.*, p. 120).

and uncorrelated components, we shall proceed to transform them still further into factors that are correlated and therefore mutually dependent.

The search for 'primary abilities,' defined in purely mathematical terms—the non-negative factors of Thurstone, or the non-fractional factors of Stephenson—seems to me entirely illusory, if defined as the final goal of factor-analysis. Such a quest is as needless, and as meaningless, as the search for absolute position and absolute motion in physics. It should be a cardinal axiom of empirical psychology, as it is of physics since the advent of relativity, that anything that is measurable must necessarily have the nature of a relation: and, once this axiom is accepted, we shall be quite content to say that a mental factor merely specifies a system of relations. The parallel may be pushed further. In the physical world the nearest approach to a real entity is Action (energy integrated through time). Similarly for the empirical psychologist, the nearest approach to a real entity is not the individual soul, nor yet mental energy or mental power conceived as residing in an individual brain: it is, so to speak, a man's successive performances or behaviour as integrated throughout the duration of his life, or, for practical purposes, a brief definable sample of it.

In my view, the special value of a factor in psychology is that it enables us to hold together in thought a definite but complex pattern of characteristics. To resolve such a complex pattern into a simple causal entity is therefore to forgo the chief advantages of the concept. The ordinary mind loves to reduce patterns to single atom-like existents—to treat memory as an elementary faculty lodged in a phrenological organ, to squeeze all consciousness into the pineal gland, to call a dozen different complaints rheumatic and regard them all as the effect of a specific germ, to declare that strength resides in the hair or in the blood, to treat beauty as an elementary quality that can be laid on like so much varnish. But the whole trend of current science is to seek its unifying principles, not in simple unitary causes, but in the system or structural pattern as such. Heat is no longer thought of as causing the molecular motions of a heated body; it is a mode of motion. Matter is no longer

conceived as moulded into sporadic masses whose forces introduce irregularities into the movements of other masses in the neighbouring spatial field; the irregularities in the spatial field *are* matter. The same change of standpoint, the same enlargement of our 'scope of apprehension' till it can think in terms of complex patterns instead of isolated units, is required in psychology; for there, as elsewhere, what the observer must look for and judge are not the spots of paint, but the picture—the canvas behind the picture remaining for ever concealed from his vision.

I conclude, therefore, that in describing mental life the psychologist is driven to the same position as the physicist has reached in investigating the material world.[1] Once we have left the field of applied psychology for that of psychology as a pure but empirical science, we are no longer justified in assuming a universe of individual 'objects' or stimuli acting on an individual 'subject' or mind. The objects or stimuli prove to be merely *Gestalten*; so are the minds; so is the experience which the minds have of the objects. We are reduced to the study of a changing structure of relations linking two sets of systems; and these systems themselves are for the scientist simply structures of relations.

Of isolated minds, then, in and for themselves, the empirical psychologist can know nothing; equally he can do nothing with isolated test-stimuli such as the early psychophysicists supposed they were using; and even the relations between minds and test-stimuli can be described in their formal aspect only. Here it is that we find at once the reason and the justification for using that peculiar mode of mathematical analysis which the psychologist, like the physicist, has recently been led to adopt. This is why, as I have already hinted, we have been gradually driven to measure both stimuli and reactions not by single figures, but by matrices of figures: for a matrix is essentially a device for

[1] This position had already been reached before the advent of relativity and quantum-theory had led to its general acceptance: e.g. by Russell (*Principles of Mathematics*, 1902, I, p. 468): "The only relevant function of a material point is to establish a correlation between all moments of time and certain points of space. . . . Thus a material point may be replaced by a many-one relation."

describing a pattern or a system of relations, without specifying what is the nature of the terms related. To the contemporary psychologist who believes that conscious phenomena are best described in terms of the patterns or *Gestalten* that they display, rather than in terms of atomistic sensations held together by 'laws of association,' matrix algebra should obviously appear the appropriate quantitative tool.

The Ultimate Inadequacy of all Purely Quantitative Statement.—Yet even matrix algebra has limitations which in the end may prove an encumbrance. The introduction of numerical quantities, though scarcely avoidable in applied psychology, raises difficulties in pure psychology, and even in applied psychology may lead us astray unless supplemented by other modes of statement. It is these difficulties, as we have already seen, that excite the protests of the intuitionist school. What schedule of measurements, they ask, can possibly describe the mind of a Leonardo or a Cézanne? How can we assess their unique accomplishments as the sum of so many units, or multiply this set of achievements by that set of figures so as to obtain a true weighted average? Even the substitution of a matrix for a single figure does not dispose of the objection; for in matrix algebra we are still condemned to cross-multiply and add: and every form of factor-analysis hitherto employed by psychologists has treated all operations as reducible to summation, weighted or otherwise. Yet why should we always proceed by addition? Suppose for the sake of argument that (as the factorist assumes) a child's performance at English composition can be expressed as the 'product' of two 'factors'—his general intelligence, say, and his verbal ability—why not take the two terms literally, and find the 'product' of the 'factors'? Why not *multiply* his performances together instead of *summing* them? To which, of course, another critic may retort, why proceed by multiplication? Why not square, or take higher powers still? Why not, in short, seek some subtler mathematical function, more complex, and therefore more elastic, and so capable of supplying, if necessary, *different* expressions for different mental processes?

To these questions several answers may be offered. The first, and perhaps the most final, were it only convincing, would be one we have already encountered, namely, that the simplest functions have the highest *a priori* probability. Indeed (though their views are never expressed precisely in these terms), that would seem to be the answer favoured by the majority of those writers who incline towards Spearman's methods or towards methods developed out of his—for example, Thurstone and Thouless.[1] " The desire to find ' realities ' behind the phenomena," says Thomson, " appears to be strong in Thurstone : his belief that, when ' simple structure ' is achieved, the factors have more significance than that which attaches to mere coefficients is of the greatest interest," and seems to imply a " refreshing " faith that " mathematical elegance is bound to correspond to physical or mental entities or actualities." [2] The same attitude, too, would, I suppose, be taken by those mathematical physicists whose familiar ' principle of simplicity ' or ' overdeterminism ' Thurstone takes to be axiomatic for every branch of science. " Of the laws that fit the data," we are told, " the simplest is most likely to be correct " [3] ; and " of all functions by which our laws can be expressed, linear functions are the simplest." If we presented our tables to a worker of experience in the field of applied mathematics, without stating that they were based on psychological rather than physical measurements, he would assuredly advise us to seek the simplest formula ; and if he gave any reason (other than the mere saving of labour) it would no doubt be that " the simplicity of a formula is a better guarantee of probability than accuracy of fit." [4] Yet, when dealing with the complex phenomena of psychology as distinct from the simple phenomena of physics, I myself find it difficult, as I have already confessed, to believe that the simplest explanations can always claim the highest *a priori* probability.

[1] Cf. the discussion of much the same issue between Thomson and Thouless, *Human Factor*, IX, 1935, p. 3.

[2] *Loc. cit.*, 1938, pp. 4–5.

[3] Jeffreys, H., *Scientific Inference*, 1931, p. 38.

[4] Jeffreys, *loc. cit.*, p. 40 ; cf. Johnson, *Logic*, Pt. II, chap. xi, p. 248.

A second, and to my mind a more plausible, reply is to admit that, in an intricate field like that of human measurement, we must be content with approximations, and accept the simplest approximations as the safest, though not perhaps the surest, until we have concrete evidence to show what form the further complications can be assumed to take. Thus, the more cautious factorist can start by postulating *a priori* an unknown analytic function in the most general form — $g = f(x_1, x_2, \ldots, x_n)$; he can then defend himself for choosing linear or additive functions to begin with, by claiming that the unknown function may be expanded by Taylor's theorem, and that, for a preliminary estimate, the first term in the expansion alone need be retained.[1]

Such a way of bridging the gap seems satisfactory enough so long as we assume that a quantitative formulation of some kind or other is admissible; it certainly seems sufficient for all practical purposes. For theoretical purposes, however, we ought, I think, frankly to recognize that the question—which mathematical function is "most likely to be correct"?—is really meaningless in psychology, because psychological phenomena, so far as we know, do not strictly obey any of the familiar mathematical laws. But there is an alternative. Our difficulty is this. Ideally, in our generalized arguments, it is not sufficient to substitute symbols to represent the unknown *variables*: we need further symbols to denote unknown functional *relations*. This suggests that the best tool for factor-analysis will not be quantitative at all: it will be a development of symbolic logic [2] rather than of

[1] Cf. *Marks of Examiners*, p. 251. It is curious to discover that if we substitute multiplication for addition, many of the transformations regularly made in factor-analysis still yield the same results. This is most easily demonstrated in the case of the simple hierarchy, where the 'two-factor hypothesis' can be shown not to depend on the assumption that the relations are additive. We may even calculate factor-saturations by taking geometric means instead of arithmetic, i.e. by a multiplication method rather than by a summation (*loc. cit.*, p. 283).

[2] Just as the addition and multiplication of simple quantities are special cases of the addition and multiplication of matrices, so addition and multiplication as understood in arithmetic and algebra are really special cases of more general operations carried out by logic. It would not be difficult to generalize most of the factorial arguments along these lines. Thus, we have already seen that the 'product theorem,' on which the hierarchy is based,

algebra. And actually, as we have seen, there is now an instrument of analysis which enables us to reason with exactitude, and at the same time to avoid specifying, not only the variables, but also the relations between them. It carries the somewhat uninformative title of the ' theory of groups,' and has been defined as " a kind of super-mathematics in which the operations are as unknown as the quantities on which they operate." To parody Bertrand Russell's famous definition, we might say that it consists of sums in which the mathematician can never know what the sums are about, nor what figures he is working with, nor yet what mathematical operations he is supposed to be performing, nor even whether his operations are mathematical at all.

I have already noted the successful use of the method in problems of quantum physics. There is little question in my own mind that the theory of groups could be applied with equal success to the analogous problems in psychology ; for, if it is doubtful whether material processes are subject to the laws of addition and multiplication, the doubt is far greater when we turn to mental processes : in psychology even more than in physics,[1] " not only the actors, but even their actions are unknown." Here, then, as it appears to me, is a line of advance which the theoretical factorist might well attempt in the near future.

Practical Implications.—My emphasis on relations, however, as forming the only concepts we can safely work with, is not to be regarded as a conclusion for the theorist alone : it has implications for the practical worker as well. In the psychological clinic the notion that a child's mind is a kind of substance with causal properties of its own—' abilities ' and ' tendencies ' that can be summed up by a few simple assessments—engenders very primitive modes of examination and treatment. Teachers, parents, doctors,

is really a special case of the product theorem as relating to classes, which in turn is a special case of the product theorem as relating to proportions. An equation merely states a mutual implication. The root of an equation defines a class. Multiplication by a selective operator is equivalent to the operation of selecting a class. And so on.

[1] Eddington, *loc. cit. sup.*

and even psychologists, are still very apt to assume that what is called the ' problem child ' has only to be brought to the psychologist's consulting-room, tested, interviewed, and observed as it were *in vacuo*, and the examiner can then pronounce what is wrong with the child.

In the past this tacit assumption was responsible for at least half our failures with the neurotic and the delinquent. The ' problem ' never lies in the ' problem child ' alone: it lies always in the *relations* between that child and his environment. Neither the delinquent child, nor yet the nervous child, nor even, as a rule, the backward child, can be properly understood, unless the examining psychologist has investigated, not only the child himself, but the conditions under which he is living at home, at school, and wherever he spends his leisure hours, and so is able to gauge how the child, on the one hand, and his parents, friends, and teachers, on the other, are constantly interacting.

This seems to me to be especially important in obtaining reports on the child's character. Not only for purposes of research, but also for purposes of clinical diagnosis and vocational guidance, the common practice is to request observers to grade persons according to their supposed emotional or moral qualities. But, when we correlate such gradings, we usually find that they throw light on very little else besides the observers' own implicit views. What we ought to grade is the overt behaviour of the persons under review, not the presumable qualities of their minds; and behaviour consists essentially in relations—in relations between the person and the conditions under which he lives.

A striking example is to be found in the so-called human instincts, about which so much controversy has been waged. From the standpoint of the individual psychologist, instincts are ' factors,' not genetic factors but descriptive factors, factors in the logical rather than in the biological sense: they are " little more than convenient headings under which certain reactions to certain stimuli can be recorded." Whether we are to postulate an inherited set of quasi-reflex mechanisms to explain such ' instinctive ' reactions is a problem belonging to an entirely different

sphere of work; and, however it is solved, much the same headings, I imagine, perhaps with slightly modified labels, will in practice still be used.

A similar but more dangerous instance of such reification is the attribution of unconscious behaviour—i.e. of acts carried out without the person being aware of them—to the unconscious 'parts,' 'regions,' 'tendencies,' or 'wishes' of the mind—in short, to 'non-cognitive mental factors' still regarded as concrete entities—a notion which seems usually associated with a strong belief in physical and temperamental 'types.' Everyone who has read the case-reports compiled by the young student will have noted how a smattering of psycho-analytic theory, reinterpreted, if he is up-to-date, in quasi-factorial language, can lead to absurdly artificial explanations couched in highly speculative terms, and suggest a wholly misguided treatment.

There is, beyond question, no richer field awaiting the factor-analyst than that of so-called instinctive and unconscious behaviour. A beginning was made many years ago. But the few recent attempts to supplement psychoanalysis by factor-analysis have tended to reinforce rather than to refine the crude notions that have hitherto obtained in regard to 'Kretschmerian types,' 'Freudian factors,' and the like, in this particular field. Most medical psychologists, it is true, still bluntly profess to "believe only in clinical research, not in statistical research" (and by clinical research they seem primarily to mean drawing conclusions from one or two cases only instead of from a number). Statistical investigators have lent substance to the implied reproach by supposing that, if they take a hundred cases instead of one, they can safely substitute rough impressionistic assessments for data gleaned from methodical examination and prolonged case-histories. The two methods of approach must supplement each other; either alone will be highly equivocal. And in this branch of psychology the revised view of 'factors' that I have advocated above would, I am convinced, lead to more trustworthy descriptions of the intricate influences at work, and display clear issues for joint research.

At the same time, the attempt to apply factorial con-

clusions to practical problems brings into strong relief the helplessness of the mere factorist so long as he tries to rely on his own unsupplemented efforts. It is as though the surgeon were to trust to the study of gross anatomy alone, declining all hints from physiology or cytology. No doubt, in the history of nearly every complex science, the study of broad relations and of observable types comes first. Factorial psychology, with its correlation and classification of persons and its correlation and analysis of traits, plays much the same part in general psychology as the older 'systematic botany' and 'morphological botany' in general botany. Just as the study of plant classification and plant structure are now supplemented by 'functional botany' and 'plant genetics,' so too, as more direct experimental methods become available, the first provisional results of factorial psychology will have to be supplemented, and even very largely superseded, by the functional and genetic study of the mind. Conversely, I believe, in the other biological sciences, many of the problems which have not yet yielded to direct attack could be elucidated, and perhaps partly solved, if analysed by the statistical devices of the factorist.

The Applicability of Factor-analysis in Other Sciences.—And this brings me to the last conclusion that emerges from my inquiry. Once we discard the notion that our 'factors' are essentially 'factors in the mind,' once we realize that 'factor-analysis,' so far from being a device adapted exclusively to the problems of the psychologist, is simply a quantitative refinement of common logical procedures, we shall not only appreciate more justly its special merits and its inevitable limitations as an instrument for studying the mind: we shall also perceive its manifest applicability to other fields of work.

I have already cited a research in which factor-analysis has been successfully used to examine the existence and nature of physical as well as mental types. Its extension to the study of numerous other anthropological or ethnological problems is almost equally obvious. I will mention only one example of special interest to current psychology. The statistical study of race-differences has proceeded on the assumption that human races form clear-cut types,

adequately definable by averages, with diagnostic characters in perfect correlation, as though human races had remained isolated and unmixed, like the varieties and the subspecies of wild animals. On the contrary, throughout prehistoric and historic times, human groups have freely migrated and freely interbred, in a way no other creature has ever done; in consequence, as recent genetic principles have forced us to recognize, the notion that human beings are still classifiable into a few racial types, with little or no overlapping, becomes wholly untenable. Accordingly, as I have elsewhere suggested, the statistical issues can no longer be dealt with by the mere comparison of averages, irrespective of variation and correlation, but must be attacked by the analysis of variance and covariance—in short, by a method which regards the so-called races as fluctuating combinations of genes, resulting in relatively stable patterns of characteristics, each pattern definable by a factor.[1]

The differences between different social classes or groups might also be studied along similar lines. Here I am thinking, not so much of alleged innate differences in intelligence or temperament, but rather of effective differences in attitudes, preferences, or beliefs, as they influence the actions of different sections of the community in our social, economic, and political life. Investigations on such problems must necessarily be planned on a large scale; and that in turn introduces difficulties both in collecting and in analysing the data. So far, the main field for extensive surveys in social psychology has been the elementary school. The wireless offers yet another easy avenue for gathering rough facts on an exceedingly large scale. The

[1] A preliminary trial of methods has been made by several of our students. J. I. Cohen has obtained data for Jewish and non-Jewish persons, though his figures and his inferences are perhaps open to some criticism (cf. 'Determinants of Physique,' *J. Mental Science*, May 1938, and Ph.D. thesis, University of London Library). P. C. Hu has made a comparative study of Chinese, English, and Anglo-Chinese children. Miss M. Davidson and others have applied factorial methods to test-results obtained from English and Welsh students and children. It is too early to pass any judgment on the value of the conclusions emerging; but the applicability of the procedure would seem at any rate to be demonstrated.

inquiry on artistic taste (alluded to above), which was conducted with the aid of the *Listener* and a series of wireless talks, was partly intended to test the possibility of social inquiries of this kind; and I think it may be said that the outcome at least showed the suggestion was feasible. In their formal character, it will be observed, all such investigations are essentially researches on the existence and nature of social 'types' and 'type-factors.' [1]

[1] In an early memorandum that I was asked to submit, when a psychological committee was formed in connection with the B.B.C., I suggested that tastes and preferences of radio listeners might be studied by sampling methods along the lines previously used in psychological and educational surveys. For the opportunity of collecting data on artistic taste by a direct appeal to the listening public, I am indebted to my friends, Mr. Charles Siepmann (formerly Director of Programme Planning at the B.B.C.) and Miss Margaret Bulley (who assisted in preparing the test material): a preliminary analysis of the results will be found in her book *Have You Good Taste?*

A somewhat similar survey had been carried out for the L.C.C. on children's tastes and preferences for films at the cinema; and, at the request of certain firms, analogous methods were used later on in an endeavour to ascertain the varying attitudes of particular types of customers towards different types of goods, wrappings, and advertisements. On problems of this kind a large amount of work has been carried out in America, but rarely with a factorial technique. G. Gallup has more recently familiarized us with the notion of sampling popular views in the surveys carried out by the American Institute of Public Opinion (for a description of the methods employed, see Katz and Cantril, *Sociometry*, I, pp. 155–79): it would be highly instructive to plan one or two inquiries of this sort with a more adequate sampling and factorial technique.

In all such surveys a number of interesting theoretical issues are involved to which the factorist might well turn his attention. The most important are the nature of the best method of sampling and of the best method of weighting. In this country such problems have chiefly arisen in connexion with investigations for education authorities. Thus, in an early survey of the abilities and attainments of the London school population, it was obviously impracticable to carry out a complete or exhaustive survey (like the census), or even a survey by simple or random sampling (conforming to the requirements of the statistical textbook). Accordingly, a twofold procedure was attempted which should combine the merits of an intensive 'complete' survey with those of an extensive 'representative' survey. The 'representative' principle meant the sampling of schools of certain 'types' from certain social strata or districts, and an effort by factorial methods to ascertain the best weighting for each. Now that education authorities are beginning to ask for psychological assistance in conducting such reviews, it is essential that the psychologist should be ready to outline an adequate technique. My own view is that, in general, the existence of such social, economic, or local 'types'

Another field in which, I venture to suggest, a factorial technique would be most fruitful is that of medicine. In a recent publication [114] I have endeavoured to point out how such a procedure may elucidate the classification and diagnosis of nervous disorders, and of mental subnormalities generally, among school children; and, by means of data collected with the help of school medical officers and others, we have been hoping to show how it may prove equally fertile in the study of physical disease. Other research-students, specially qualified in the relevant field of work, have been experimenting, so far with decidedly promising results, to demonstrate its adaptability to problems of industry, economics, and plant-fertility.[1] Indeed, the type of problem for which factor-analysis would seem the most appropriate method is one which, so far from being confined to psychology, is common to all complex sciences where work is in a preliminary stage. In biology, in medicine, in agriculture, in meteorology, in almost every sphere of research where we are dealing no longer with a few simple conditions operating on a few large bodies in the cosmic void, but with clusters of interacting causes, affecting highly composite reagents, there, as I have more than once ventured to contend,[2] innumerable questions are waiting to be solved, or at least unravelled, by the factorial methods that the psychologist has evolved.

can best be established by the 'analysis of variance' (which I regard as essentially a factorial method); but so far the method has rarely been tested in actual practice.

For an account of the aims and methods of social surveys, cf. A. F. Wells, *The Local Social Survey in Great Britain*, 1935, and *id. ap.*, F. C. Bartlett *et al.*, *The Study of Society*, 1939. The surveys carried out by economists, though far more adequately planned than those of sociologists, have been chiefly limited to economic conditions; but their statistical methods, supplemented by analysis on factorial lines, might serve as models for similar surveys on social and psychological problems.

[1] In these fields the most interesting outcome would seem to be a demonstration of the value of applying factor-analysis to problems hitherto treated by the analysis of variance, just as it appears to be equally valuable in the psychological field to apply analysis of variance to problems hitherto treated by factorial methods.

[2] [93], p. 313; *Nature*, cxliv, 1939, p. 533.

CHAPTER VIII

SUMMARY AND CONCLUSIONS

To reduce my chief conclusions to a few brief statements will be by no means easy, since the conclusions themselves are necessarily so tentative. My primary object has been to determine a little more exactly the nature of so-called mental 'factors' by examining the form of proof by which those 'factors' are established. I conclude that factors as such are only statistical abstractions, not concrete entities. To resolve a test-performance into '*g*' and '*s*' no more demonstrates the existence of a general and a specific 'ability' than describing a breeze as north-west implies the combination of two currents from separate quarters of the sky. We use factors in psychology as we use rectangular co-ordinates in other sciences, not because we believe that the phenomena investigated are necessarily dependent upon a few, isolated, independent causal agencies, which operate in the mind or brain, and which a sound method of factor-analysis should successfully isolate, but merely because such simplified descriptions enable us to organize our facts into a more logical system and help us to state our inductive arguments more cogently, and so endow our predictions with higher probabilities.

Mathematically, a factor is simply an average—usually a weighted average—of certain measurements empirically obtained. Logically, it is simply a principle of classification —a principle by which both tests (or traits) and the persons tested may be classified. Four kinds of factors may be formally distinguished—(i) general, (ii) group or bipolar, (iii) specific, and (iv) error factors, i.e. those possessed by all the traits, by some of the traits, by one trait always, or by one trait on the occasion of its measurement only. This fourfold division broadly corresponds with the traditional fourfold scheme of predicables, viz. *genus*,

species (or *differentiæ*), *proprium*, and *accidens*; and might be regarded as a quantitative adaptation for the case of *variables* of a qualitative scheme originally developed for the case of *attributes*. More particularly, the 'bipolar factors' resulting from general-factor analysis correspond to classification by progressive dichotomy, while the 'group-factors' resulting from group-factor analysis provide an equivalent classification of the same phenomena by co-ordinate classes.

From the four-factor theorem (as it may be termed) all the familiar factor theories may be derived. The differences, however, between the several kinds of factor are not absolute, as these theories commonly assume, but merely relative; so that what is a group (or bipolar) factor in one table may emerge as a general factor in another or a specific factor in a third.

The primary value of such factors must obviously consist in their utility for purposes of systematized description. Whether or not any factor actually extracted or computed happens to have a psychological significance is a problem that must depend, not on the method of factor-analysis employed, but upon the proper and relevant selection of traits and persons. Factors cannot be invoked for purposes of inference or prediction without additional data and additional assumptions not included in the table of data factorized or in the conclusions immediately drawn from it. Indeed, the inferential and predictive use of factors as such is far more limited than is commonly supposed; and its validity must rest upon a due observance of the general conditions of all inductive inference—in particular, on the stability of the results from one set of observations to another and on certain *a priori* postulates about the subject-matter of the inference, postulates which are seldom explicitly announced in factorial work. Thus, the significance and the reliability of the conclusions deduced will turn quite as much upon the design of the experiment by which the data are secured as upon the technique of the analysis by which the data are examined. By itself factor-analysis can at most describe only the general structure of the mind or of the population: functional problems require

other methods of research, which may in turn illuminate or modify the provisional concepts reached by mere factorization.

So far as it seeks to be strictly scientific, psychology must beware of supposing that these principles of classification can forthwith be treated as ' factors in the mind,' e.g. as ' primary abilities ' or as ' mental powers ' or ' energies.' Factors specify not unitary qualities but systematic patterns; not active entities, but relations between what we loosely call the mind and what we vaguely call its environment; i.e. they specify systems of relations between two sets of relational systems.

These views, I have shown, appear at once more obvious and more plausible when we consider, not only the more usual kind of factorial work that begins by correlating traits, but also the complementary mode of approach that begins by correlating persons. And the whole interpretation, I believe, is closely in keeping with the *Gestalt*-like conceptions that the modern physicist offers us when he describes the material world and that the metaphysician has from time to time put forward in attempting to describe reality. On the other hand, the current treatment of factors as causal abilities implies an antiquated attitude towards both scientific and metaphysical issues.

As a method of inquiry factor-analysis reveals a close and suggestive analogy with the mathematical methods employed in modern physics. It might, therefore, be still further refined by adopting or adapting some of the newer instruments of analysis that have been successfully employed in that sphere. In particular, the use of the theory of groups might obviate many of the objections commonly urged against the crudities of mere quantitative calculations when applied to the mind. Finally, it is argued, instead of the psychologist invoking postulates and principles appropriate only to the simpler sciences, the more complex sciences might in their turn profitably borrow factorial methods from psychology.

PART II

THE RELATIONS BETWEEN DIFFERENT METHODS OF FACTOR-ANALYSIS

CHAPTER IX

CLASSIFICATION OF METHODS

Problem.—From a mathematical standpoint, as we have seen, the aim of factor-analysis is, broadly speaking, to find a reversible (or ' non-singular ') transformation which will enable us to reduce a given matrix of figures, such as a table of test-measurements or of correlations, to an equivalent form which shall be of the simplest possible type, at least so far as such automatic simplification is consistent with the problem and scope of the research in hand. To effect such a transformation various working procedures are available, which are to be found in textbooks on higher algebra or analytic geometry[1]; and, corresponding to these procedures, different methods of analysis have been proposed by psychologists during the last ten or fifteen years—the ' two-factor method,' the ' centroid method,' the ' method of principal components,' the ' method of trigonometrical rotation '—each having its own special merits and its own group of advocates.

The sponsors of any one method warmly criticize the rest. Thurstone considers Hotelling's results to be " psychologically meaningless "; Kelley pronounces Thurstone's method to be " logically unsound "; Spearman rejects all three in favour of his own original procedure (with simple extensions where necessary); and his criticisms have been supported by Stephenson in this country and Holzinger in America. The general student thus comes to regard each method as wholly incompatible with the others, and is at a loss to know which to choose.

In a *Memorandum* on factorial methods, drawn up at

[1] Cf. [26], [44], [73], [74]. For the practical computer perhaps the clearest exposition of the classical devices of the professional mathematician is to be found in the recent work on *Elementary Matrices* by Frazer, Duncan, and Collar (Cambridge University Press, 1938, esp. chap. iv).

the request of the English Committee of the International Institute Examinations Inquiry, I ventured to maintain that the results of the different procedures, though often discrepant at first sight, could nevertheless be regarded either as approximations to, or as linear transformations of, one and the same set of theoretical values : so that, in spite of each author's rejection of every rival method, a direct reconciliation seemed always possible. My review, published (with some abridgment) as an appendix to the Committee's report [93], was concerned primarily to describe and briefly prove the formulæ available, not to examine their inter-relations in detail ; and, naturally enough, the assertion of their essential agreement has been freely called in question. Formal demonstrations of the equivalences would have been too long and technical to embody in the report itself ; but they have since been printed elsewhere—usually as incidental to some specific piece of research, and not always, it would seem, in an accessible and easily intelligible form. In this paper I propose to bring together, in a way that may be readily understood by the non-mathematical reader, what appear to be the most important of these results.

The Definition of Factors.—To begin with, it is essential to understand a little more precisely what we mean by a factor. As we have already observed, hardly any factorist has offered a satisfactory formal definition. But the discussion in the preceding paper should at least indicate what form that definition must take. Factors as such must be defined, not by trying to identify them with concrete causal entities in an objective world of individual minds or nervous systems, but by *specifying the operations by which they are to be obtained.*[1] The actual practice of

[1] Writers who insist on defining factors " by our concrete psychological needs " have protested against the " artificiality " of my " purely mathematical mode of definition." In reply, may I remind them that the principle here adopted has become almost universal in quantitative science ? " It has come to be the practice in introducing physical quantities that they shall be regarded as defined by the series of measuring operations and calculations of which they are the result : those who associate with the result a mental picture of some entity disporting itself in a metaphysical realm of existence do so at their own risk " (Eddington, *loc. cit.*, p. 71).

factorists, so far as they share a common procedure, would seem to suggest the following formulation.

(i) *Components and Factors Generally.*—Let there be n observable variables measured for N individual cases. The $n \times N$ empirical measurements so obtained— $\{m_1, m_2, \ldots, m_n\}$ say—will in general be correlated with one another in a more or less arbitrary fashion. We can, however, assume that it is always possible to express them exactly or approximately in terms of r new hypothetical variables, $\{p_1, p_2, \ldots, p_r\}$ say, $(r \lessgtr n)$, which will be related to each other in some simpler, specifiable way : so that we can regard the empirical set of measurements as functions of the latter, and write

$$\{m_1, m_2, \ldots, m_n\} = f_i \{p_1, p_2, \ldots, p_r\} \; (i = 1, 2, \ldots, n).$$

Then these r hypothetical variables $\{p_1, p_2, \ldots, p_r\}$ may be called ' factors,' or more accurately ' components,' of the observed measurements ; and any particular system of components will be defined by the system of functions f_i, or the inverse of that system if it has one.

Such a transformation can be effected in an infinity of ways ; and in carrying it out, it is, as a rule, implied that the new variables will be so chosen that they can serve as reference values. Thus, (i) usually the specifiable relations between them will be the simplest possible, namely, those of independence or non-correlation ; (ii) usually, too, their number will be the fewest possible, or at any rate they will be fewer[1] than the original n variables, if only because

[1] There are many important exceptions to this rule, which, however, would seem to be apparent exceptions only. Thus, Garnett ([37], 1919, p. 346) takes $r \geq n$, envisaging one ' general factor,' n specific factors, and a varying but presumably small number of group-factors. This is virtually the view of Thurstone (whose proofs resemble Garnett's in many ways) except that he would not specify a general factor as distinct from the other common factors, and always insists that the number of common factors must be less than n. Garnett also mentions the possibility that the one general factor (and the other factors, if desired) may be so transformed that r may be any number whatever : this is presumably to allow for Thomson's sampling theory, where the number of ' elements ' is assumed to be indefinitely large and the number of ' factors ' a maximum (viz. 2^{n-1} : see below, p. 294, footnote). Spearman, on the other hand, always takes $r = n + 1$. Hotelling

none but the statistically significant components will actually be calculated ; so that a comparatively small number of the new components will account for far more of the observed variance than would be accounted for by an equal number of the original variables ; (iii) nearly always (though not in my view necessarily [1]) the function f, will be taken to have the simplest possible form, i.e. to be a linear function, so that we can write $M = FP$, where P is (by i) an orthogonal matrix (or horizontal section of such a matrix) and F a linear operator or matrix pre-multiplier. It might seem reasonable to add, since reference-values are usually needed for *permanent* reference : (iv) the set of components selected must be stable for all samples of the categories from which the n and N items are drawn, and for the same samples on different occasions. That, however, is a result which would require separate demonstration for each set, and is, as a matter of fact, hardly ever explicitly established.[2]

and Kelley put $r = n$ precisely. However, each of these writers recognizes that all but a few of the numerous components deducible in theory are in practice likely to be devoid of statistical significance.

[1] Just as we should not assume *a priori* that both height and weight can be expressed as linear functions of the same factor or set of factors, so, it seems to me, we should not assume that mental performances must necessarily be treated as linear functions of the related factors. It is interesting to note that a good many of the familiar factor theorems do, as a matter of fact, hold good with more complex functions ; and, in any case, a linear function may be regarded as supplying a first approximation to the more complex function (cf. above, p. 241). Nor is it, I think, commonly realized that the product-moment formula, on which our correlations and regressions are regularly based, is merely the simplest case of a more general formula. Thus, in the numerator we may write the generalized covariance $\frac{1}{n}\Sigma x^p y^q$, where p and q do not necessarily $= 1$, with corresponding expressions for the two generalized variances in the denominator. The system based on these three parameters should, in theory, enable us to specify the law of dependence between x and y, no matter how complex it may be.

[2] It might perhaps be contended that we only factorize tests or trait-measurements that are known to be 'reliable' ; so that the reliability coefficients obtained (*a*) from different sections of the same test, and (*b*) from applications of the same test at different times, would guarantee the stability of the initial variables. Yet that hardly suffices to prove the stability of the factorial pattern. It would be better to include the separate applications or sections as separate variables in the correlation table, and so show that the

Accordingly, it does not seem proper to include this requirement in the formal definition.

This fourth requirement would mean, in the case of tests and persons,[1] (*a*) that the factor-saturations for a particular test must not change, when new tests are added to the old battery of tests; (*b*) that the factor-measurements for a particular person must not change, when new personsare added to the original sample of the tested population; and perhaps (*c*) that if the same tests are applied to the same persons at different times (e.g. to the same children at different ages or later on when grown up) both the factor-saturations and the factor-measurements should remain unaltered. It is only necessary for the last point to be stated for us to see how precarious it is. Yet, unless it is assumed to hold good at least in some degree, the study of mental development generally, and predictions about the subsequent development of the individual in particular, become very difficult.

It is curious that, while nearly all factorists insist that any 'factor' to deserve the name must be stable, nevertheless, hardly any of them formally prove this stability. Few seem to have fully realized either the advantages or the limitations such stability confers. One great advantage would be that the indeterminacy, which is bound to haunt us so long as we try to deduce a plausible set of factors from a *single* table of figures only, would be greatly diminished, and often wholly abolished, if we insisted that our factors should be demonstrable, not in one table, but in a succession of tables—tables obtained with varying batteries of tests and with different groups of individuals. We believe in a general cognitive factor (*g*), not because it was conclusively established by a single research, but because it appears and reappears in almost every collection of cognitive tests; and, as a test of identity (by no means the only test), we carry one and the same test, or group of tests, (the Binet tests or some recognized series of written intelligence tests) from one research to another. Instead of rotating a factor pattern by trial and error until it satisfied some *a priori* scheme, we ought to rotate it mechanically until it fitted two or more tables. Indeed, the device of rotation, as ordinarily applied, is an informal

common factors cover them. However, the most effective procedure is to prove, by an appropriate criterion, that the factors obtained from different investigations remain approximately the same (cf. p. 41). With but *one* investigation our sole resource is to attempt an 'efficient' estimate of the population parameters by the method of 'maximum likelihood' [50].

[1] And, of course, within the margin allowed by the sampling error.

test of stability: only then it tests, not so much the psychological significance of some particular factorial pattern in the light of our preconceived notions, but the stability of our preconceived notions about the factors in the light of yet another empirical table.

On the other hand, as I have argued elsewhere, a rigid adherence to the postulate of invariance would force the factorist to surrender (*a*) the idea of specific factors, peculiar to a single test or a single person, (*b*) the idea that the factors obtained by correlating persons are essentially different from those that are obtained by correlating tests, and ultimately, I believe, even (*c*) the idea of standardizing scores on the basis of the given sample. This postulate, therefore, rules out so many ideas still widely favoured by factorists that it seems hardly proper to include it at the very outset in the formal definition of factors in general.

To begin with, therefore, it will be best to introduce only the first three of the four restrictions mentioned above. And it would perhaps be convenient to keep the term 'factors' for a particular and determinate set of components, expressly selected so as to conform to these conditions, and to use the term 'components' for 'factors' in any broader sense.

(ii) *Orthogonal Linear Factors.*—If we accept these three conditions, we can take a further step. For simplicity, let us keep chiefly to the type of problem for which factor-analysis has been most commonly employed, namely, the testing of n traits for N persons. As a rule, the sample population tested will not only be comparatively numerous ($N > n$), but will also vary from one inquiry to another; the tests will not only be fewer in number, but, if duly standardized as regards material and procedure, will presumably remain constant. Hence it seems natural to define our factors in terms of the tests alone.

Now by matrix multiplication (a common device in matrix work for reducing a given set of figures to a simpler form) the N individuals can be at once eliminated. We have $R = MM' = FPP'F' = FF'$, where (assuming M to have been suitably standardized) R is the $n \times n$ matrix of correlations, and F is a matrix of factor-loadings of order $n \times r$. Accordingly, incorporating the restrictions stated above, we may now define our factors as *a hypothetical set of*

mutually independent, statistically significant linear components, derived from an observed set of measurements by a homogeneous linear transformation; and each particular factor may be defined by stating its correlations with the several tests as specified by the corresponding column in the matrix of the inverse transformation, namely, F. In geometrical language the resolution of the test-measurements into factorial components may thus be very simply described as consisting in a change to orthogonal reference axes.

We can put this in a more concrete way. Since $M = FP$, the jth element in the ith column of M (e.g. the jth measurement for the ith person) can be split into a sum of weighted measurements, $m_{ji} = \sum_k^r f_{jk} p_{ki} = {}_1m_{ji} + {}_2m_{ji} + \ldots + {}_rm_{ji}$ say.[1] Here the kth figure, ${}_km_{ji}$, is attributable to the kth factor only. On covariating the kth figure for all the persons, we obtain $\mathbf{f}_k\mathbf{p}_k \times \mathbf{p}_k'\mathbf{f}_k' = \mathbf{f}_k\mathbf{f}_k'$, a symmetric matrix of rank one. Thus, we can summarize our algebraic interpretation in words, and say: *a factor is the class of any set of variables, including parts of observed variables, whose covariances form a perfect hierarchy.* For example, if we imagine a set of n fictitious tests and if we suppose that the $\frac{1}{2}n(n-1)$ relations between them can all be expressed in terms of n constants only, one constant for each test, then all those tests would measure one and the same factor, i.e. they would all belong to one and the same irreducible class.

Even so, however, the operations specified by the definition are not determinate. No actual tests are likely to satisfy precisely these exceedingly simple relations; and in general, an empirical covariance matrix, R, can only be fitted exactly by a factorial matrix F containing n^2 figures. Yet, in virtue of its symmetry, R contains no more than $\frac{1}{2}n(n+1)$ different figures, and so yields at most $\frac{1}{2}n(n+1)$ equations for the purpose. If we confine ourselves to significant components only, we shall require fewer figures than n^2: but this general instruction is too vague to yield a unique procedure. Unless, therefore, an investigator tells us what particular method he is adopting to obtain determinate values for F, we cannot say precisely what he understands by a factor.

(iii) *Doubly Orthogonal Factors.*—So far I have followed the ordinary approach to factorial work. "The object of

[1] Cf. Appendix II, Table IVb.

the factorial analysis," it is said " is to find F " [1] ; and with that, as a rule, the matter usually terminates.

To the practical psychologist, concerned more with individual persons than with the abstract analysis of mind, the determination of F appears merely a means to an end, not the end itself. Hence I myself should prefer to base my definition on the equation $P = WM$, taking the rows of P to specify the new variables and the rows of W to specify the operation required. As we have just observed, it is commonly assumed that the rows of P are uncorrelated ; but the grounds that make it reasonable to secure non-correlation in the factor-measurements make it equally reasonable to secure non-correlation in the factor-loadings and regression coefficients.[2] Thus not only the rows of P but also the columns of F will be uncorrelated. This

[1] Thurstone [84], p. 70. And again : " When R is given experimentally, the problem is to find F " : I should rather say, when M is given experimentally, the problem is to find P.

[2] This is most obvious if we turn to the case of correlating not tests but persons. In such a case the factorial description of each person is given by his column of saturation coefficients ; and these, when reciprocity obtains, are identical with the factor-measurements that would be obtained for that person by correlating the tests or traits. If, therefore, with the latter procedure the factor-*measurements* are assumed to be uncorrelated, it must also be assumed that with the former procedure the factor-*saturations* will be uncorrelated.

But, apart from this special argument, we may consider the question on more general grounds. When two factors prove to have very similar saturations in the same set of tests or traits, can we still maintain that they are pure and independent factors ? Surely, factors that are described in almost the same way are almost the same factors : independent factors must have independent descriptions. The saturation coefficients for many of the rotated factors in Thurstone's recent study—his factors for verbal relations (V) and for deductive reasoning (D), for example—are positively correlated : both have high saturations for most of the ' verbal ' tests and of the tests of ' verbal reasoning ' ; is it not natural to suppose that this is because they both contain a common factor ([122], p. 115–16) ? Again, many of Stephenson's pairs of ' type-factors ' have very high negative correlations (e.g. [92], p. 302) : but the signs of a column of saturation coefficients may always be reversed ; is it not therefore natural to suggest that in such cases we have to do mainly with a single bipolar factor, and not two independent factors ? The very object of factor-analysis, we are told, is to reduce the number of variables. Why, then, choose a relatively large number of correlated factors, when a smaller number of uncorrelated factors will do the work equally well ?

condition yields another $\frac{1}{2}r(r-1)$ equations. And, assuming the test-variances which form the diagonal elements of R to be known or otherwise deducible, the procedure for finding F and W now becomes determinate; both can be expressed in terms of the *latent roots and latent vectors of the correlation matrix* R ([93], [101]).

The properties of these constants enable us to give a clear and definite meaning to the set of factors thus extracted. They will be the set of r components (where r may be any number from 1 to n) that yield the *best fit* both to the initial matrix of measurements and to the covariance (or correlation) matrix derived from them. The closeness of the approximation will be determined, as usual, by the principle of least squares. Accordingly, this method of analysis may be termed the 'least squares method.' [1]

The method is merely a special application of the principle regularly adopted in multiple correlation. n examiners (or the same examiner using n different tests) endeavour to assess the intelligence of N candidates. The first or dominant factor is then defined as that weighted combination of the test results which presumably has the highest correlation with the characteristic that they were all intended to measure. More generally, we may say that each factor in turn is determined as being that particular component which has the largest mean square correlation with all the test-measurements, or residuals of the test-measurements, still waiting to be factorized.

The guiding principle can be expressed more precisely if we revert to algebraic symbols. The procedure indicated is equivalent to seeking for M, first of all the best-fitting matrix of rank one, namely, M_1 (say) $= \mathbf{f}_1\mathbf{p}_1$. Then, if the residuals $M - M_1$ are too large to be considered statistically insignificant, we proceed to seek the best-fitting matrix for these residuals, M_2 (say). We shall then have found for M the best-fitting matrix of rank two, namely, $M_1 + M_2$. And so we continue. It follows that $M_1M_1' = \mathbf{f}_1\mathbf{f}_1'$ will also be the best-fitting matrix of rank one for the covariance matrix $R = MM'$; and similarly for the other product matrices. If at each stage we make the residual variance as *small* as possible, we shall automatically be making the factor variance as *great* as possible. Hence, for those

[1] In my first paper, using the method of 'simple summation,' I assessed the amount of discrepancy by taking the mean deviation of the hypothetical from the observed correlations, i.e. the average of the individuals regardless of signs ([16], 1909, Tables V and VI): the least-squares method amounts to assessing it by the root-mean-square deviation, which requires the method of 'weighted summation.'

who prefer the language of 'analysis of variance,' we may describe our procedure as seeking first the factor whose contribution to the total variance shall be as large as possible: then the factor whose contribution to the remaining variance shall be as large as possible; and so on: until the ultimate remainder is zero or rather statistically negligible as judged by the probable error.

It is easy to show that these requirements are equivalent to reducing the correlation matrix R to a diagonal canonical form by an orthogonal transformation. We obtain $L'RL = V$; so that we can write $F = LV^{\frac{1}{2}}$, V denoting the diagonal matrix of latent roots (or factor-variances), and L the orthogonal (or semi-orthogonal) matrix of latent vectors (or direction cosines) ([93], p. 290; [102], p. 177).

(iv) *Factors defined by Selective Operators.*—The foregoing procedure brings the analysis into line with that adopted for analogous problems in other sciences. We have $M_k = \mathbf{l}_k\mathbf{l}_k'M = E_kM$, where $E_k = \mathbf{l}_k\mathbf{l}_k'$, is a unit hierarchy, and $E_1, E_2, \ldots, E_k, \ldots$ form a 'spectral set of selective operators.'[1] Thus, we can carry our previous

[1] In its simplest application the notion of a selective operator may be explained as follows. Let M be a mixed population consisting of r mutually exclusive classes or types, say, Europeans (M_1), Indians (M_2), and Chinese (M_3). Let E_1 be the selective operator which sorts out only the Europeans; E_2 the selective operator which sorts out only the Indians; and so on. Then

$$E_1M = M_1;$$

but $E_1M + E_2M + \ldots + E_rM = M_1 + M_2 + \ldots + M_r = M;$

$$\text{i.e. } \Sigma E_k = \mathbf{1} \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

Again $E_1E_1M = E_1M_1 = M;$

hence $E_k^2 = E_k \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$

(i.e. E_k is 'idempotent'): for, if we try to select the Europeans from a selected group containing nothing but Europeans, we reach the same group as before.

Once more $E_2E_1M = E_2M_1 = 0$; hence $E_2E_1 = 0$, and generally

$$E_jE_k = 0; \ (j \neq k) \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

for if we try to select the Chinese from a selected group that contains no Chinese, we obtain an empty class.

Operators possessing these three properties are said to constitute a spectral set. (The term 'spectral' is derived from the use of such operators in atomic physics where on the basis of a spectral analysis of a mixed radiation an attempt is made to segregate 'pure components' along the lines of the classical experiment of Stern and Gerlach.)

formulation a step further, and say that, given a suitable set of observed variables, each of their r factors may be defined by a *unit hierarchy with a definite factor-variance attached.* The final arithmetical solution of the problem will be given by calculating the terms of the series

$$M_1 + M_2 + \ldots + M_r = E_1M + E_2M + \ldots + E_rM = M$$

which may be called the factorial expansion of M: or, more succinctly, by calculating the nr factor-measurements, $P = WM = V^{-1}L'M$ ([101], p. 84).[1]

In the pages that follow, I propose to show that all other current solutions may be regarded as derivatives of this more specific solution.

(v) *Oblique Factors.*—Throughout the theoretical part of this discussion I shall as a rule use the term factors to mean what are commonly called ' orthogonal ' factors, as distinct from ' oblique,' i.e. factors for which the factor-measurements are uncorrelated or independent,[2] and which can therefore be represented by rectangular axes. The relations between oblique and orthogonal factors are simple; and, for the 2- and 3-dimensional case, will be familiar from elementary geometry. (i) With *oblique* factors, the correlations between factor-measurements and test-measurements can be obtained by multiplying the matrix of factor-

[1] For working instructions and procedure, see Appendix I. The reader who is unfamiliar with the matrix notation used in the foregoing argument will follow the points more easily if he turns to the elementary example worked out in Appendix II, Tables I and IV.

[2] In describing factors as statistically independent, uncorrelated, or 'orthogonal,' most writers appear to be thinking solely of correlations between the factor-measurements, and not of correlations between the factor-saturations as well; but they do not state this explicitly. Thus, Thurstone's preliminary analysis is based on the assumption that " the factors are uncorrelated " (p. 61); but the equations expressing this assumption show that he is referring only to the rows of the factor-measurements in his ' population matrix ' (P_4). He postulates that " the n tests which constitute the battery are so selected that the columns of the factorial matrix (F) are *linearly* independent," and this is evidently meant to imply that they need not be " statistically independent " (p. 57). With my ' simple summation ' method the factor-saturations usually have a low correlation, which may be regarded as merely an effect of an imperfect mode of approximation. With the centroid-summation method as used by Thurstone (highest correlation in leading diagonal) the correlations may be appreciable.

saturations by the matrix of correlations between the several factors. When the factors are orthogonal, the latter becomes a unit matrix. Hence (ii) with *orthogonal* factors, as we have seen, the correlations between factors and tests are given directly by the factor-saturations ([101], p. 88).

An analysis which yields oblique factors may be regarded as an incomplete or partial analysis, since all correlation has not been eliminated; and the factors so obtained may be regarded as 'partly analysed' or 'mixed factors.' The two modes of analysis correspond to the two kinds of 'selective operators' which the physicist has recognized: first 'fractional' operators, which effect a partial analysis into constituents that are not pure; secondly, 'spectral' operators, which effect an analysis of a mixed aggregate into pure constituents. The difference is usually defined by saying that fractional operators, unlike spectral operators, are not idempotent ([115], p. 159).

In theoretical inquiries the advantage of working with independent factors will be sufficiently obvious from analogies with partial differentiation and partial correlation. It is, however, part of an ultimate logical requirement:[1] all reasoning about complex variations becomes, not only simpler, but more rigorous, if we can reduce it to terms of concepts of which any one can be taken to vary, while the others remain constant. In factor-analysis we are enabled to reduce the number of effective elements in the covariance matrix from $\frac{1}{2}n(n+1)$ to n; and the terms eliminated (the covariances as distinct from the variances) are precisely those which are most difficult to manipulate algebraically, to compute arithmetically, and to estimate statistically. In short, with independent variables, we can treat the variances as simply additive.

In practical applications such orthogonal factors may not always be necessary or even desirable. Eventually, perhaps, we may be able to identify certain of our so-called factors with well-marked physiological agencies or mechanisms, e.g. the secretions of certain glands, or 'unit characters' in the genetic sense: but the activities of the various glands will almost certainly show some low degree of correlation; and some of the unit characters will be, not independent, but linked. Meanwhile, if we regard our hypothetical factors

[1] On the logical importance of independence in its various forms—notional, connectional, implicational, and the like—see Johnson, *Logic*, pp. 54 f., 108 f., 219 f.

merely as empirical principles of classification, then, though logical rigour requires such principles to be independent, practical exigencies will still impose a compromise. The classifications of the school psychologist, for example, have often to be made in terms of partly correlated factors. When he explains that his theoretical factors classify children according to verbal, arithmetical, and manual abilities or disabilities, teachers inevitably assume that he is referring to the more concrete classifications of pupils into schools or forms according to what he would regard as 'mixed factors,' not 'pure' or independent. Rarely if ever does the teacher think of differentiating (say) between a child whose disability in verbal subjects is due mainly to his inferiority in general intelligence and one whose disability is due to inferiority in what may be loosely termed pure verbal capacity, i.e. an inferiority which affects solely his power to understand or manipulate words.

In my educational work, therefore, I have found it necessary to distinguish between what I have called 'mixed' or 'joint' factors (e.g. 'concrete abilities') and 'independent' factors (e.g. 'pure' or abstract capacities), or sometimes (adopting the phraseology of Binet) 'compound' and 'partial aptitudes.'[1] Other writers have recognized a similar distinction. Alexander, for example, uses the terms 'functional abilities' and 'independent factors' to express it. "The so-called functional abilities," he writes, "are not independent traits; they are resultants. . . . Verbal and practical abilities can be resolved into three (independent) factors—*g* and *v*, and *g* and *F*" ([82], pp. 117 *et seq.*). Rotation will often recombine independent factors into what Thurstone calls 'oblique reference vectors' ([84], p. 154). Eysenck, for instance, has endeavoured to show that in Thurstone's recent work on *Primary Abilities* many of the rotated factors, e.g. *V*, *P*, and *W* (verbal relations, perceptual functions, and verbal fluency [122], Table 4) are positively correlated, and that each includes a common factor, similar to Spearman's *g*, in addition to some pure or independent factor peculiar to itself (cf. [133], p. 272). The repeated application of Spearman's two-factor procedure to submatrices of residuals successively obtained from the same initial table of correlations has also led in the past to factors which would seem to be correlated rather than independent. From

[1] *L'Année Psychologique*, XIV, p. 32. Cf. also *The Backward Child*, pp. 459 *et seq.* In my earlier *Notes*, the physicist's terms 'spectral analysis' and 'fractional analysis' were used to indicate the evident analogy, but (as indicated below) it would seem a little unwise to introduce into psychological nomenclature a rather puzzling pair of terms for a distinction that is already designated in half a dozen different ways.

a mathematical standpoint such modes of analysis appear partial or incomplete, like a fractional distillation that only partly separates the various ingredients.[1]

Methods Available.—The various methods proposed need not be described afresh. Formal descriptions of the main types (taken from roneo'd notes compiled originally for the use of my own research-students and based on early articles by the chief authors concerned) were included in my *Memorandum* [41]; but the systematic expositions since published by Thurstone [84], Kelley [85], and Holzinger [106], all of them remarkably lucid and suggestive, render my previous account of their principles not only

[1] A clear and systematic account of the relations between the two types of factors has been recently given by Holzinger and Harman (*J. Educ. Psych.*, 1937, XXVIII, pp. 321–45). More recently still, in his contribution to the discussion on 'Factor Analysis' Stephenson has developed the same distinction in a suggestive way ([137], pp. 100–2). Here, and in a personal communication, he criticizes my own account in two respects that deserve a fuller reply than was possible at the Symposium ([137], pp. 92). In the first place, though he emphasizes the distinction between 'concrete,' 'functional,' or 'unanalysed abilities' and such abstract or hypothetical factors as Spearman's 'general factor' (*g*) and my 'verbal factor' (*v*), he rejects "the supposed 'independent factors' as gratuitous assumptions." Factors like *g* and *v* he now prefers to call 'fractional factors,' to distinguish them from the 'non-fractional factors' that play a central part in his theory of Q-technique ([138], p. 272). Yet on a later page he explains that these 'fractional factors' are uncorrelated, and are obtained by analysing the concrete or 'non-fractional abilities.' Surely in that case they are identical with what Alexander and I understand by 'independent factors,' so that after all these latter cannot be 'gratuitous assumptions.'

Secondly, he adopts the algebraic argument from the article of mine just quoted ([115], pp. 157–8); but thinks it should be applied, not, as I applied it, to show that the factors complying with it cannot be further analysed, but to show what conditions the 'unanalysed abilities' must fulfil. This would make his unanalysed abilities or 'non-fractional factors' identical with Holzinger's original 'bifactors,' i.e. bifid factors ([106], p. 6): for in both cases the saturation coefficients for the two 'pure' factors are assumed to be proportional. In an empirical matrix, however, such a precise proportionality is so improbable that Holzinger has now given up the conception ([107], p. 53). In the fuller manuscript version of his paper Stephenson points out that his non-fractional factors are analogous to the factors derived by Tryon with the aid of 'cluster-analysis'; and in his more recent work (particularly on types with 'Q-technique') he deliberately avoids a 'general factor,' so that his non-fractional factors are, as it were, group-factors

needless but somewhat out-of-date.[1] My indebtedness to earlier writers still, particularly to Spearman (whose brilliant work has after all inspired, directly or indirectly, the numerous alternative methods put forward to supplement or supersede it), to Godfrey Thomson, William Brown, and Hotelling, will be obvious. Nor, except for a few incidental comments, shall I attempt to weigh the advantages of the different procedures. All the controversies of the past appear to have overlooked one indisputable fact, namely, that there is no one royal approach, superior to all others and suitable for use on all occasions. Each particular class of problem demands its own peculiar

derived directly from the empirical correlations without the elimination of the first or dominant factor : in certain respects, therefore, they are analogous to the factors obtained by Thurstone in his work on *Primary Abilities* after rotation (cf. [133], p. 272), and depart from those that would be obtained by a 'Spearman analysis.' His own description, however, brings his present methods more nearly into line with those of Tryon than with submatrix methods, as used by previous factorists (e.g. by Holzinger, by Stephenson in his earlier articles, and by myself). Indeed, from recent correspondence with him I gather that the conception of 'non-fractional factors' was partly devised to cover what Tryon has called 'operational psychological unities' : (the instances given by both authors are much the same ; while Stephenson's 'fractional factors' cover what Tryon terms 'radical and orthometric components' : cf. Tryon, [86], [131]).

A third criticism has been urged against his statement, namely, that his extension of the term 'fractional' from methods of analysis to the resulting factors is in conflict with the usage of previous writers and with his own. In [138], p. 277, he refers to "fractional factors, which will usually be *oblique*." But this would seem to be a slip of wording. The context makes it plain that he is referring to a "fractional *analysis*, the results of which will usually be oblique factors." In his last paper he makes this clear, since he describes the analysis of correlations by multiple-factor technique as leading to 'fractional *orthogonal* factors' ([136], p. 242). The proposed extension is quite consistent, if a little confusing : there is no inconsistency in maintaining that a 'complete *analysis*' should yield a 'fractional' *factor* and that a 'fractional' (i.e. partial) *analysis* should leave us with a 'non-fractional' *factor* (i.e. one that is imperfectly analysed into its ultimate parts).

[1] As has already been explained (pp. ix–x), this account was written before the publication of Prof. Thomson's book (*The Factorial Analysis of Human Ability*) which gives an admirably clear and impartial account of all the chief methods available, and forms by far the best introduction to the whole subject. Actual methods of calculation are described and illustrated in Appendix I, Tables I-XI, pp. 449–86 below.

devices. For our present purpose it will be sufficient to group the various methods according to procedure rather than aim. The distinctions I shall observe are set out in the table below. There will, of course, always be combined and transitional types; but that does not invalidate the main dividing lines.

TABLE I.—*Classification of Procedures*

I. Analysis of Variance
II. Factor-analysis of
  A. Covariances } between
  B. Correlations }
    1. Persons } by
    2. Traits }
  (*a*) Group-factor Methods, with
    (i) Non-overlapping Group-factors
    (ii) Overlapping Group-factors
  (*b*) General-factor Methods, with
    (i) Simple Summation
    (ii) Weighted Summation

CHAPTER X

VARIANCE, COVARIANCE, AND CORRELATION

A. *Variance*

I SHALL begin with a method that is not usually classed as a method of factor-analysis. In the *Memorandum* just cited it was pointed out that many of the problems which psychologists have been accustomed to attack by means of factor-analysis might be more satisfactorily solved by the procedure described by Fisher and others under the title of 'analysis of variance.' An instance of its use will be found in chapter xxi (Table VII). Since the method has been but little employed in psychology,[1] I shall begin by

[1] The most obvious problems for which the analysis of variance would seem to be the appropriate method are those of mental and scholastic testing; but we have also found it of special value in analysing the results of laboratory experiments, since there the sample is almost inevitably small. In some of our early investigations, the method was tentatively applied to data obtained from junior county scholarship and other examinations, and appeared to confirm the results obtained by 'correlating tests'; more recently we have applied it to back mark-lists for the Teachers' Diploma and Certificate examinations. A still more convincing comparison has since been completed by Miss Cast, using marks obtained from an artificial examination expressly designed for the purpose ([134], pp. 257–69). Similar comparisons have also been attempted with data obtained by ranking (e.g. ranking picture postcards in tests of artistic appreciation); here the results of analysing variance were compared with those of 'correlating persons.' Experiments with the method are thus by no means numerous: but in every case they have been sufficiently successful to show that the proposal is something more than a mere suggestion.

It may be added that many of the early investigations carried out by psychologists with ranking methods, along the lines first suggested by J. McK. Cattell, were in principle analyses of variance, with the mean variation substituted for the standard deviation or its square: (cf., for example, F. L. Wells' analysis of the 'Variability of Individual Judgment,' *ap. Essays in Honour of William James*, 1908, pp. 509–50). Wells tabulates some of his initial data in full: this makes it possible to re-analyse his figures by more modern methods—e.g. by correlating persons, and so to demonstrate the

attempting to indicate what are the general relations between the two procedures. So far as I am aware this has never yet been done.

The similarity both in object and in method is somewhat obscured by the current terminology. What is usually described as an ' analysis of variance ' is really an analysis of a matrix of measurements by means of its ' square-sums.' Similarly, what is commonly regarded as a factor-analysis of correlations is, in my view at any rate, really an analysis of the same matrix of measurements (or a standardized version of it) by means of its ' product-sums ' ; and the factorization consists essentially in ' rotating axes ' until the product-sums are reduced to an equivalent set of ' square-sums ' ([101], p. 77). Under simple conditions the relations between the two methods are elementary and direct, and can be expressed in algebraic form.

Let us confine ourselves first of all to the analysis of correlations or covariances between tests. We may suppose that k tests have been applied to N persons [1] and that the test-measurements are expressed as deviations about the average for each test. The simplest case is that in which (i) the initial measurements are in standard measure

inadequacy of the earlier statistical procedure. Mr. Eysenck has carried this out at my suggestion ; and has shown that, contrary to Wells' own inferences, his table reveals a positive and significant general factor. This is also confirmed by a formal analysis of variance. Such results appear to me to be a striking illustration of the valuable advance in method made during the last 30 years.)

[1] It is not easy to find a notation which shall be consistent with that customarily used in factor-analysis and in the analysis of variance ; indeed, in expositions of the latter the usage of different writers varies in a way that is most confusing to the student. I suggest that where (as is usually the case) we are concerned with testing a classification by one criterion only (*either* by persons *or* by tests) we follow Fisher and Yule, and write k for the number *in* each of the ' classes ' (or ' populations ') whose means are to be compared, retaining N or n to denote the number *of* ' classes,' according as the ' class ' denotes the set of measurements for the same test or the set of measurements for the same person. I shall use Q (*quadrata*) to denote square-sums, P to denote product-sums, and V to denote the mean square or variance obtained by dividing the square-sums by the number of degrees of freedom. The present use of the symbols P, Q, V, and F has no relation to their use to denote certain matrices or axes in factorial theory.

(according to the common practice in psychological testing) and (ii) the general factor is obtained by the simple-summation method, with the 'reliability' assumed to be perfect. These assumptions will simplify the exposition, but are not essential to the argument: (for more general purposes it would be better to assume that the variances may differ, so that, according to their differing complexity or their different relation to a common factor, the several tests are differently weighted when they are summed).

The primary object of factor-analysis is then to demonstrate the 'existence' of a single significant source of variation, influencing all the test-measurements in the matrix to be analysed, namely the first 'common factor.'[1] The factor is specified by the hypothetical factor-measurements for the N persons, the measurements being computed as weighted sums or averages of the k tests: (the weighting, however, makes little difference to the figures finally accepted; and, except so far as it is needed to eliminate gross and arbitrary differences in the scales of the several tests, is hardly ever carried out in practice). Thus the problem of factor-analysis is equivalent to showing that the differences between these factor-measurements are significant.

If there were only two persons, we could apply the usual test for the difference between their respective means, by taking the ratio of the difference observed to the estimated standard deviation. But when the number of persons is large, the task of testing the significance of all the possible differences separately is greatly magnified. It is, however, possible to devise a single critical ratio, and so judge their significance by a single test. This is the principle adopted in the analysis of variance. Let x_{ji} denote the measurement for the ith person in the jth test;

[1] In 'general-factor' analysis, this can be extended to the case of multiple factors, since each successive factor is taken in turn, and the same calculation is begun over again with a fresh matrix of measurements, namely, the table of residuals: in each case we look for a single factor pervading the entire table in question. In 'group-factor' analysis we look for a factor common to a sub-matrix only, i.e. that part of the table of residuals which relates to a particular sub-group of tests.

let $\sum_{j}^{k}\sum_{i}^{N} \bar{x}_i^2 = k\sum_{i}^{N} \bar{x}_i^2 = Q_m$; $\sum_{j}^{k}\sum_{i}^{N} (x_{ji} - \bar{x}_i)^2 = Q_r$; and $\sum_{j}^{k}\sum_{i}^{N} x_{ji}^2 = Q_t$: so that $Q_t = Q_m + Q_r$. Finally, let

$$F = \frac{(Nk)_f}{(N)_f} \cdot \frac{Q_m}{Q_r} = \frac{V_m}{V_r},$$

where $(Nk)_f$ and $(N)_f$ denote the 'number of degrees of freedom'[1] on which the estimated variances $V_m = Q_m/(N)_f$ and $V_r = Q_r/(Nk)_f$ are based. Then F is the critical ratio. Tables showing the probability that a given value of F (or of $z = \frac{1}{2}\log_e F$) may be reached or exceeded with $(N)_f$ and $(Nk)_f$ degrees of freedom are given in the text-books of Snedecor and Fisher [50].

If we regard all the marks obtained by a single person as forming an array or a class, then the F-ratio appears as an elaboration of a ratio that is far more familiar to the student of psychology—the correlation ratio[2] η. The relation between the two is given by $\eta_{ij}^2 = \frac{Q_m}{Q_t} = \frac{(N)_f \cdot F}{(Nk)_f + (N)_f \cdot F}$. It will be noted that Q_t is the value

[1] The 'number of degrees of freedom' is found by subtracting from the number of measurements made the number of 'constraints' or independent linear relations obtaining between them. Here this means in effect that we compute the 'mean' by dividing the corresponding sum, not by the number of measurements summed, but by the number of *independent* comparisons that can be made with a sample of that size. When the sample contains more than 30, we can average in the more familiar way: e.g. we can take the residual variance to be Q/Nk simply. Strictly, however, we should remember that the residuals are deviations about the means of their respective columns. Hence only $N(k-1)$ are free to vary independently: in each of the N columns the kth figure can be deduced from the rest. If the means of the rows are also zero, one of the figures in each row can be deduced from the rest. Hence the degrees of freedom are still further reduced to $(N-1)(k-1)$. If the standard deviations of each row are also equalized to each other and to unity, yet a further limitation is imposed: and so on, with the calculation of each additional parameter. It is the difference between these two methods of estimation that gives rise to the difference between the biased and the unbiased estimate of the intra-class correlation, referred to below.

[2] It should be noted that, since j (here symbolizing the jth test) does not enter into the formula for calculating η_{ij}^2, the so-called 'correlation' for which η_{ij} is the corresponding 'ratio' does not *necessarily* indicate a correlation of anything with j as a quantitative variable or rank in a fixed order, though in the past the correlation ratio has generally been employed to measure such a correlation.

taken by Q_m when Q_r is zero, that is, the maximum value of Q_m. This gives an alternative method of determining Q_t which is often convenient.

If we proceed to correlate the several tests, we can estimate the extent to which they agree by the average intercorrelation, $\bar{r} = \frac{1}{k(k-1)} \Sigma r_{ii'}$ $(i \neq i')$, i.e. by the mean of the correlations of each test with every other, excluding all self-correlations. (This device has more frequently been used in the past where not tests, but persons, or persons' estimates of traits, have been correlated.)

If the tests are all in standard measure, it is not necessary to calculate each intercorrelation separately; for when the means and the standard deviations are equal, the average intercorrelation is identical with what is called in the analysis of variance the 'intra-class correlation' (r_{int}), that is, the correlation calculated by disregarding the allocation of measurements to definite tests and taking all possible pairs of measurements for one and the same individual.[1] Shortened methods of computation are thus available. We have in fact $\bar{r} = r_{int} = \frac{k\eta^2 - 1}{k-1}$ (an extremely useful formula which we can call the 'corrected square-sum ratio').[2] Where the initial measure-

[1] Fisher, *loc. cit.*, pp. 198 *et seq.* Yule and Kendall, *loc. cit.*, pp. 254 *et seq.*

[2] The equation given in the text corresponds with the formula usually given for calculating the intra-class correlation (cf. Yule [110], p. 255; Fisher [50], pp. 202, 211). This, however, yields a biased estimate, which is in general somewhat too small; from the equivalences given by Fisher ([50], p. 213) a more accurate estimate can be obtained by taking into account the limitations in the degrees of freedom. The relation between the two may be exhibited as follows:

$$\text{(i) biased } r_{int} = \frac{(k-1)Q_m - Q_r}{(k-1)Q_m + (k-1)Q_r} = \frac{kQ_m - Q_t}{(k-1)Q_t}$$

$$\text{(ii) unbiased } r_{int} = \frac{n(k-1)Q_m - (n-1)Q_r}{n(k-1)Q_m + (n-1)(k-1)Q_r} = \frac{V_m - V_r}{V_m + (k-1)V_r}$$

In either case the second expression is the most convenient for computation.

The use of the average intercorrelation as a form of 'saturation coefficient' and of the grand average of all the intercorrelations as an indication of the 'factor-variance' was an early device (e.g. *Mental and Scholastic Tests*, 1921,

ments have the form of ranks, it affords the simplest and speediest way of estimating the predominance of a single general factor. It takes but a fraction of the time required for a formal analysis carried out in the usual way.[1]

Let us now suppose that the correlations between the several pairs of tests are separately calculated in the usual way, and that a full factor-analysis is then attempted. For simplicity, we may keep to the method of simple summation, as adopted by Thurstone and others, and assume that the test-measurements, as usual, are in standard measure.[2] The calculated self-correlations will then be unity; to simplify the proof, we can retain these values since, as Thurstone observes, "the diagonal entries (in the correlation matrix) may be given any value between zero and unity without affecting the results markedly, especially when the number of variables is large" ([122], p. 108). With the simple-summation method, we begin by calculating the sum of all the (inter-class) correlations, including the assumed self-correlations. Let us write, for the average of the complete set of inter-class correlations calculated in this way, $\bar{r}_c = \frac{1}{k^2} \Sigma r_{ii'} (i, i' = 1, \ldots k)$. Then, the average saturation coefficient for the general factor will be

$$\frac{1}{k} \Sigma r_{ig} = \sqrt{\bar{r}_c} = \eta = \sqrt{\frac{Q_m}{Q_t}}.$$

pp. 136–7; Thomson and Bailes, *Forum of Education*, IV, 1926, pp. 85–96). The substitution of the average correlation for a correlation with an average is (as we shall see later on) merely a corollary to the important theorem that the covariance between sums is equal to the sum of covariances.

[1] Burt, *loc. cit.*, p. 17. It was used with ranks by Pelling, Dewar, Wood, and myself in preliminary studies of the general æsthetic factor by correlating persons: with ranks the denominator for η^2 can, of course, be directly determined by the formula used in the calculation of a rank correlation, viz.

$$Q_t = \frac{k(k^2 - 1)}{12}.$$

[2] If we remove these two limitations, and covariate test-measurements with a differential weighting, extrinsically or intrinsically deduced, and then employ the method of weighted summation, the general argument will be the same: but somewhat elaborate complications will be introduced owing to the new relations set up between what at the outset were independent variables.

The relation between the two average correlations (with and without self-correlations) is shown by

$$\bar{r}_c = \frac{1 + (k-1)\bar{r}}{k} = \eta^2.$$

It will be remarked that in this case the use of the correlation ratio is equivalent to assuming known self-correlations. It is for this reason that the intra-class correlation or its equivalent, which (as we have seen) does not involve these latter, may be regarded as an improvement on, or correction of, the bare correlation ratio. Evidently the critical ratio, given above for testing significance, if computed from the saturation coefficients, could be expressed as follows :

$$F = \frac{(Nk)_f}{(N)_f} \cdot \frac{(\Sigma r_{ig})^2}{k^2 - (\Sigma r_{ig})^2}.$$

Generalizing this result, we see that factor-analysis, like analysis of variance, may be regarded as a division of the total test-variance (or of a simple function of it) into two parts : one part due to a common source of variation and the other a residual variance. And our conclusions may be summed up as follows. Under the simple conditions assumed, (i) the average saturation coefficient, calculated from the completed correlation table, is identical with the correlation ratio, i.e. with the ratio of the observed standard deviation of the means to the maximum standard deviation ; (ii) the average intercorrelation (i.e. excluding the assumed self-correlations) is identical with the intra-class correlation, which may be regarded (by those who prefer to think in terms of the more familiar expressions) as a kind of corrected correlation ratio ; (iii) in factor-analysis the chief comparisons turn on the mean of the squares of the saturation coefficients ; in analysis of variance it turns on the square of the mean of the saturation coefficients.

So far we have assumed only one criterion of classification. We have considered the various test-measurements as grouped by persons only. In such a case we inquire whether the persons differ significantly as regards their average performances in each of the k tests ; and each person's average is taken as indicating his ' factor-measurement ' in the ' general factor for the tests,' i.e. in the general

quality which all the tests are presumed more or less to measure. But frequently it is possible to cross-classify or sub-classify one and the same set of measurements by two or more criteria simultaneously. In such cases the analysis of variance goes on to split the residual variance (or rather residual square-sums) into two or more portions, much as it began by splitting the total. The 'analysis of variance with two criteria' corresponds to those factorial inquiries in which we are interested both in the general factor for tests and in the general factor for persons: with two criteria, therefore, the problem is to demonstrate the significance of the differences between the means *both* for the columns (persons) *and* for the rows (tests) of the measurement matrix. When these two sources of variation have been removed, we are left with a doubly centred bipolar table of 'discrepance,' the means of both rows and columns all equal to zero, analogous to the doubly centred bipolar table of residual measurements and residual correlations obtained in factor-analysis after the general factor or factors have been removed. In the analysis of variance these residual figures are utilized for the study of what is called 'interaction,' that is to say, of the differential relations between the items in the rows and in the columns. This corresponds to the study of secondary factors in multiple-factor analysis. In the analysis of variance, as the term 'interaction' implies, its reciprocity is taken for granted. When these further developments are borne in mind it becomes obvious that there must be a large variety of psychological problems which could be attacked with profit by these newer statistical methods.[1]

[1] In planning experiments for the subsequent analysis of variance, there is one important device which has proved of great use in agricultural research and which might also be applied with advantage for numerous psychological investigations, namely, the 'randomization' of the arrangement of the selected items. We have found this principle of special value in overcoming more effectively many of the familiar difficulties in experiments on the efficiency of different methods of training, testing, and the like (e.g. more particularly in ruling out the irrelevant effect of the lapse of time, difference in difficulty of test-materials, etc.); it also enables us to obtain valid results from smaller groups and often to solve several questions simultaneously: (an illustrative research by W. D. Seymour—on the influence of visual adaptation on speed of reading with different persons—will shortly be published). It is suggestive to note that Fisher, in his recent book on the *Design of Experiments*, actually chooses a psycho-physical experiment to introduce the principles he has elaborated, and to illustrate the need for randomization ([109], pp. 13 f.).

Interaction, it should be added, may be of various kinds. That noted in the text corresponds to the interaction of the items in the rows with the

When persons are correlated instead of tests, much the same argument can be used. But a further problem then arises, because, though we may nearly always assume a normal distribution for persons, we cannot always do so for tests or test-items. Very frequently—for instance, in the investigations on æsthetic appreciation already mentioned—the test-items are chosen so that there are approximately equal differences between successive pairs, and are ranked by the persons in serial order.[1] In such cases it would appear, from a comparison of calculations actually carried out, that, unless N and k are small, the same equation may be still used for F (with the necessary modifications in the degrees of freedom) and (for approximate estimates) the same criterion may still be employed. When N and k are small, the obvious alternative is to calculate the actual distribution and use χ^2 or some related function.[2]

The advantage of using analysis of variance lies not only in the more precise tests of significance that the method allows (a point of special importance with small samples), but also in the further possibilities which the method opens up for the effective planning of experiments. On the other hand, factor-analysis, if somewhat more laborious to carry out, claims to yield additional information which could not be reached by an analysis of variance alone—e.g. the detailed factor-saturations and weighted factor-measurements. Moreover, in applying analysis of variance to psychological data, there are, it will be seen, two assumptions involved which factor-analysis may at first sight seem to avoid: first, analysis of variance assumes (or appears to assume) that the

items in the columns (e.g. in agricultural experiments, the interaction of different types or treatments of soil with different types or varieties of plants). There may, however, also be an interaction between the row items (or the column items) among themselves (e.g. of the different treatments with each other). In a psychological research, this would take the form of interactions between the different persons assessed, or again between the different traits assessed. Here there seems to be a new field of research which hitherto has been almost entirely neglected and yet which is of supreme importance in a complex subject like psychology, where almost every measureable characteristic is apt actively to influence every other.

[1] Cf. Dewar [118], p. 33. A fuller comparison of the results of factor-analysis and analysis of variance for problems in which persons rather than tests are correlated will be found in the sequel to B. M. D. Cast's paper [134].

[2] Friedmann, 'Use of Ranks to Avoid Assumption of Normality Implicit in the Analysis of Variance,' *J. Amer. Stat. Assoc.*, XXXII, 1937, pp. 675 *et seq.*

variance to be analysed is a definite and absolute quantity ; secondly, it assumes (or appears to assume) that the means of the arrays or classes—i.e. the factor-measurements—must be simple unweighted means. I shall touch upon these difficulties later on in another connexion. Here it will be sufficient to refer the reader to the few comparative studies in which both methods have been used with the same data and which show by concrete example how far such difficulties can be met in special cases.[1]

B. *Covariance*

The more familiar methods of analysis that we have now to compare were intended primarily to be applied to tables of correlation. Nevertheless, all of them may be used equally well with tables of covariances (i.e. with averaged but unstandardized product-moments about the means), or, for that matter, with tables of unaveraged product-moments or even of unadjusted product-moments (i.e. product-moments about an absolute or arbitrary zero that is not the arithmetical mean). The employment of covariance instead of correlations has both its merits and its limitations. The arguments in favour of it have been set forth elsewhere; and need not be repeated here.[2] The objections demand examination at somewhat greater length.

They have been most clearly summarized by Stephenson during a recent discussion on 'Statistical Methods in Psychology.'[3] Here he criticizes my proposal to use the 'analysis of variance and covariance' on the ground that

[1] Cf. Cast, *loc. cit. sup.* and references.

[2] Cf. above, p. 41 ; also [117], p. 419, and [93], p. 247. The chief argument, it may be remembered, was that the matrix to be analysed is the matrix of initial measurements, not (as is usually supposed) the matrix of correlations : the analysis of correlations is only a means to that end, not an end in itself. In psychology, as elsewhere, the interpretation of correlation coefficients—as distinct from regression coefficients, which are rarely calculated by psychologists—is often extremely dubious : yet there is a widespread tendency to credit them with an objective existence and a stability which, from their very nature, they cannot legitimately claim.

[3] *Proc. Roy. Soc.*, Ser. B, CXXV, pp. 415 *et seq.*

" variance in psychology is purely an arbitrary matter, depending solely on the whim of the psychologist." " Factor-analysis," he contends, " is impossible unless all measurements are first reduced to standard measure; and this is automatically effected by the usual product-moment correlation " : in short, " the only safe assumption is that the true variances are all equal for all abilities " ([117], p. 423 ; cf. [96], p. 199).

That mental measurement is to a large extent arbitrary I should never wish to deny : indeed, my proposal was accompanied by an explicit reservation, added on that very ground ([117], p. 419).[1] But arbitrariness is not fatal. The units employed for physical measurement were once as arbitrary and variable as those now employed in mental testing ; and most of them are still defined by an artificial convention. The foot, the span, the cubit, and the ell were based on the notion that each man could be trusted to provide his own standard, namely, the length of some member of his body : Henry I of England sought to eliminate variation by stretching out his own arm : (the modern physicist prefers a bar of platinum, but even he has to specify its temperature) ; David of Scotland came nearer to the psychologist's procedure, when he ordained that an inch or ' thumb ' should be an average of three thumbs—those of " an merkle man, an man of measurable stature, and a lytell man." Variance depends in part upon the way we select the population to be measured : let us

[1] Thus, in dealing with marks for the same set of scripts, where the variances seemed to depend very largely ' on the whim of the examiners,' I expressly used ' correlations between persons' instead of ' covariances between persons ' ([93], pp. 267-9). Here my procedure was criticized by another colleague on the opposite ground, namely, that the differences in the standard deviation ought not to be disregarded. Once more I agree with the reason advanced, but not with the conclusion drawn. In such cases my proposal is not (as was wrongly supposed) that the standard deviations should be ignored altogether, but that their implications should be considered as a separate problem. Let me add that I take my ' reciprocity principle ' and ' symmetry criterion ' to apply primarily to *covariance* matrices, or to correlation matrices regarded as covariance matrices. If we start by assuming those principles to be correct, and seek diagonal elements that conform to them, each principle yields an additional method for discovering or checking the most appropriate values for the variances of the items ' correlated.'

then follow King David, and define what we are to regard as a typical or representative sample. Variance also depends on what tests we choose: let us then lay down certain types of test as standards, specifying if necessary place, time, material, and the like, just as we do in defining weight or length. Sooner or later, no doubt, a regulation from the Board or Education will provide an official definition of the mental year, just as it once defined educational attainments in terms of 'standards,' and just as an Act of Parliament has defined the meaning of the yard and pound.[1]

[1] More recently Stephenson has argued that "Burt's reciprocity procedure" (using covariances between persons to obtain the same factors as were formally obtained by correlating tests or traits) "is a notable advance in technique but is insecurely founded, because he has as yet put forward no generally acceptable system of units, which is surely of first importance in any science" ([137], p. 95). This seems to overlook the fact that from 1913 to 1927 the majority of my investigations were bound up with an attempt to "put forward a generally acceptable system of units" in terms of which intellectual, emotional, and (more particularly) educational differences could be measured. Various possibilities were proposed, since for different purposes different units seemed to be required. For example, with school children it was suggested that, for rough practical purposes, the notion of the mental year as unit could be extended from intelligence to educational and emotional characteristics: such a method at once indicates wide differences in the variance of different characteristics, usually corresponding with the complexity of each (cf. [35], p. 25, [41], pp. 258 *et seq.*). For more exact measurements of general ability, the idea of 'internally graded tests' was advocated. Unlike the Binet scale, each sub-test was to consist of a graded series of homogeneous items increasing in difficulty or equally spaced, the equality of the unit-intervals being determined either by the number of children passing the items or by the number of persons judging the intervals between the items to be equal or different: the unit-intervals in the *different* sub-tests were then to be related by means of regression coefficients. Such a method of construction was slow and cumbersome, but seemed essential for exact theoretical work ([27], pp. 46, 151, and later L.C.C. *Reports*, e.g. [41], p. 138). For more general purposes still a 'standard scale,' based on the normal frequency curve, was drawn up ([35], p. 49) and widely used: it has, as a matter of fact, been adopted by Stephenson himself in some of his earlier experiments on correlating persons. With this procedure, it may be noted, if elementary aspects of a complex process are marked in standard measure and then summed (according to the 'analytic method of marking'), the total process will be automatically weighted in accordance with its assumed complexity. The last of my books opens with an entire chapter devoted to the 'Choice of Units' in mental measurement (*Backward Child*,

Already, indeed, as we have seen in a preceding chapter, many standardized tests are available for which the differences in variance are by no means entirely arbitrary. In some cases the differences can be directly estimated; in other cases they can be indirectly inferred. It then becomes evident that variance, so far from being dependent solely on the psychologist's whim, and so far from being equal for all abilities, varies widely from one ability to another. Nor are the reasons far to seek.

For example, the measurable variance of simpler processes proves usually to be appreciably smaller than that of more complex which include or incorporate the simpler, and the variance of processes that have become habitual is nearly always smaller than that of processes that are relatively novel or unlearned. Where variance cannot itself be directly measured, we still find that the simpler and more automatic processes show the smallest correlation with one another and therefore with the general factor: if we interpret the correlation as essentially a covariance, the peculiarity can at once be explained on the same lines as before. Such indications, could they be more extensively confirmed, would seem to have an intimate bearing on the procedure to be followed in extracting factors, and on the interpretation of those factors (particularly the general cognitive factor or *g*) when extracted.

In early papers on intelligence tests I drew attention to the fact that the correlation of such tests with the general cognitive factor—'intelligence,' as it is popularly termed—appears to vary very closely with the complexity of the particular process tested, and inversely with the degree to which it has been rendered mechanical or automatic.[1] Thus, when speed tests are used, McDougall's dotting test (tapping combined with aiming) shows a much higher saturation with intelligence than simple tapping; card-sorting

pp. 15–36). No doubt, as there indicated, the theoretical problem bristles with difficulties; and the units there proposed are practical devices rather than a final system. But neither that nor their arbitrariness entitles us to assume that differences in variance cannot be measured, at least with a rough but reasonable approximation (cf. also pp. 128 f., on validating such units).

[1] "The greater the complexity and the greater the novelty involved in the task, the greater also (*ceteris paribus*) is the intelligence of the performer" (*Brit. J. Psych.*, III, 1909, p. 169). "The more complex the mental process involved and the higher the mental level tested, the more completely do the test-results correspond with estimated intelligence" (*J. Exp. Ped.*, I, 1911, p. 95).

than simple card dealing; and compound reactions generally than simple reactions.[1] Similarly, with scholastic tests composition shows a higher correlation with the general factor than spelling, and spelling than handwriting; again 'problem arithmetic' shows a higher correlation than tests of the 'fundamental rules,' long division than short division, and short division than subtraction ([35], Table XVIII, p. 52). Now in each of these cases the compound processes just mentioned include the simpler. Hence it is almost inevitable that the variances of the former should be wider than the variances of the latter. This is borne out by the experimental data. For example, with card-dealing the range was 26 secs.; with card-sorting 53 secs. with the same number of cards ([17], p. 172, Table VIII). In the reaction-time experiments the standard deviations of a group of students were as follows: simple reaction, 0·08 sec., choice 0·25 sec., discrimination 0·56 sec., association 0·87 sec.

[1] The reaction-times were obtained with a d'Arsonval chronometer. The details of the work were partly suggested by experiments on Donders' 'addition theory of composite reaction-times.' With 'sensory tests' as distinct from 'motor tests' the same principle seemed to operate, but was less easy to establish. I noted, however, that in certain cases "tests of sense-perception may also be made more complex; and, when this is done, the 'intelligence coefficient' is again generally increased" (Burt and Moore, *J. Exp. Ped.*, I, 'Mental Differences between the Sexes,' pp. 247–8; the detailed results are discussed more fully in Mr. Moore's thesis). With group tests of intelligence, dealing with 'higher mental levels,' similar tendencies were again observed: thus, within the same age-group, the higher and more complex relational tests (e.g. analogies and reasoning tests) displayed much larger mean variations than simpler relational tests (e.g. opposites), just as these showed larger mean variations than the simple sensory and motor tests. There was therefore greater overlapping in the former cases between the different age-groups; and the overlapping increased with increasing correlation with intelligence. Much the same conclusion was drawn in comparing the overlap, not of age-groups, but of the two sexes: there, too, "the more complex the process tested, the wider the range of individual variation, so that, on the highest levels of all, the individual differences practically swamp the group differences."

No teacher, I imagine, would subscribe to the view that the variance for all school subjects is the same—e.g. that the individual differences between pupils in accuracy of spelling show as wide a range as their differences in English composition. In such judgments the implied unit is the least perceptible difference. As we have already seen (p. 134), such a unit can be systematically used for measuring qualitative differences in children's performances (e.g. in their handwriting, drawings, handwork, and the like); and, when this device is adopted, the increase of variation with the complexity of the test is once again very plainly shown.

With scholastic tests, the average standard deviations with children aged 12–14 (in mental years) were—for writing, 1·11; for spelling, 1·26; for composition, 1·60; thus, in terms of variance the figure for composition is more than double that of writing. Much the same holds with tests for arithmetic and for manual subjects.[1]

These and many other lines of evidence might be cited to show that "different mental processes must differ widely in their variance" and that "variance depends largely, if not mainly, on the complexity of the process measured." It was, indeed, facts of this nature that led me to suggest what I have variously called the 'complexity or integrative theory' of intelligence—the theory that the general factor measured by intelligence tests may be regarded as "an inborn capacity of the nervous system for complex organization"[2]—and seemed to justify a mode of factor-analysis which treated the standard deviation of each test as equal to its saturation coefficient for the general cognitive factor.

I do not, however, suggest for a moment that this is the whole source of differences in variance, or that the more complex mental processes are constructed by merely adding simpler processes together according to rules of arithmetic. On the contrary, I believe that the analysis of variance should enable us to detect other sources of variance besides those that are commonly admitted, and

[1] *Mental and Scholastic Tests*, pp. 402, Tables XLVI *et seq.*; cf. also [35], pp. 69 *et seq.*

[2] 'The Experimental Study of General Intelligence,' *Child Study*, IV, 1911, p. 23. In recent discussions on the 'psychometric' *versus* the 'intuitionist or clinical' approach there has been much criticism of the notion that individual variations in compound or complex processes can be treated as the additive result of variations in certain simple or elementary processes (cf. Cattell, 'Measurement versus Intuition in Applied Psychology,' *Character and Personality*, VI, pp. 114–131, and Vernon, *loc. cit.*, pp. 99–113). Those of us who have tried to combine the 'psychometric' approach with the 'clinical' have laid ourselves open to attack from both sides. To avoid such misconceptions, I now prefer to call my theory an 'integrative theory' of intelligence (cf. [16], p. 169), since compound reactions (as I have always tried to insist) must be regarded as a result, not of the *mere* addition or superposition of the simpler component reactions, but rather of their organization or integration into systematic wholes. It was for this reason that in the later paper I argued that intelligence should be described as a 'complex *synthetic* activity,' and should not be thought of as depending on the mere degree of complexity alone: (cf. *J. Exp. Ped.*, 1912, *loc. cit.*, p. 12, where it was shown that tests involving integration as well as increased complexity yield far higher correlations than the tests whose chief characteristic was an increase in complexity alone, just as these had been shown to yield higher correlations than the simpler tests; see also p. 217 above).

that the study of what (in the analysis of variance) is called 'interaction' would indicate how mental processes modify one another when they are combined. Moreover, almost from the start, increasing evidence for supplementary factors of a more specific kind indicated that a slightly better estimate for the test-variance would be obtained by adding a small and varying amount for so-called group or specific factors. This means that, in factorizing a covariance or correlation table, we ought rather to identify the test-variance with what I have called the 'complete communality,' i.e. the sum of the squares of the saturations for *all* factors. If we increase the number of our tests indefinitely, without appreciably increasing the number of factors, this figure is identical with the *limit of the multiple covariance* (i.e. of the square of the multiple correlation of the given test with all the rest). However, as we shall see later on, except with small tables, the contribution to the variance made by non-general factors is relatively slight.

Of the better-known methods of factor-analysis, Kelley's [85] is the only one that is expressed primarily in terms of covariances.[1] At the same time, however, he appears to assume that the factor loadings or saturation coefficients so derived will be precisely the same as those that would be obtained from correlations. This assumption, as I have endeavoured to prove, is not strictly true ([102], p. 193).

The relations between the results of the two procedures may be indicated briefly as follows. As we have observed, in its initial steps the problem of factor-analysis can be most generally conceived as a problem of linear dependence or rank, rather than of statistical dependence or correlation, which is only a special case of the former ([101], p. 69); the principle underlying the correlationist's procedure

[1] Hotelling refers to the "replacement of correlations by covariances," but decides "not to treat of this obvious generalization except to discuss a suitable criterion for weighting" ([79], p. 422). The special assumptions that he makes for the purpose of this criterion would, as he points out, define 'a natural unit of measure for each variate' (p. 510). Though expressed in different terms, the equation that he then reaches for the test-variances (eq. 51) would seem to be identical with my own (eq. 12, [101], p. 75). If so, my 'natural units' would be the same as his. The 'metric' adopted in the body of his paper, however, is "based on the assumption that the essential quantity to be analysed is the unweighted sum of the variances, where the total variance of each test is taken as unity," which would be (as he points out) incompatible with his equation 51 (p. 421).

is then seen to be a very ancient one, most familiar from its occurrence in solving normal equations deduced from the requirement of least squares (cf. p. 148). A matrix of order $n \times N$, such as is obtained by testing N persons with n tests, has a rank of n only, if $n < N$, and is therefore deducible from n factors, since any $(N - n)$ of its columns are linearly dependent on the rest. Self-multiplication is the obvious device for reducing such a matrix to a simpler and symmetrical shape, with an order equal to its rank; and correlation is essentially a process of self-multiplication, which (as we shall see) is reapplied in factor-analysis. But in correlating tests, before multiplying the rows, (i) we first subtract their averages: this yields, after cross-multiplication, a matrix of unaveraged covariances, but eliminates one important source of variance—roughly identifiable with the 'general factor for persons.' (Alternatively, if we start by correlating persons, we virtually eliminate the 'general factor for tests.') (ii) Secondly, we divide the product-sums by the appropriate standard deviations, which means pre-multiplying and post-multiplying the product-matrix by a non-singular diagonal matrix. This leaves the rank and the number of factors unchanged,[1] but alters the weighting of each, with the result that the relative size of one or more of the factors may at times be so diminished that it appears in effect suppressed. But, however great the apparent change, *a factorial matrix fitting the correlations can be always derived from that obtained on analysing the covariances (or vice versa) by simply prefixing the diagonal matrix (or its inverse)* ([102], p. 193, [114], p. 174). Analogous changes, it may be added, are also introduced by selection, e.g. by taking either a group homogeneous as regards the general factor, or a group differing in heterogeneity for the different correlated traits; evidently this will have the effect either of reducing the differences between the averages or of altering the magnitudes of the standard deviations.

There has been considerable disagreement between psychologists about the need for correcting correlations for attenuation before their factorization is attempted. Spearman and Hotelling, for example, would first correct; Thurstone, Kelley, and I myself[2] use

[1] Unless a further modification is introduced, it also puts the variance of each test (its self-correlation or 'reliability coefficient') equal to 1·00. The result is to throw part of the specific and error factors into the general factors. If, however, we substitute the amount of variance attributable to the significant factors only, the rank of the matrix is at once reduced (cf. pp. 154 f.).

[2] In my view, if the reliability coefficient is so low that the raw correlation requires correction, that is a reason, not for factorizing corrected correlations but for improving the experimental technique.

the raw correlations as they stand. The relation between the sets of factors obtained from corrected and uncorrected correlations is similar to that between sets of factors derived from covariances and the corresponding correlations (i.e. covariances corrected for difference in standard deviation). Assuming that Spearman's simple correction formula ([47], p. 204, eq. 155a) has been used, the factors obtained by analysing corrected correlations can be deduced directly from those obtained by analysing the raw correlations by prefixing a diagonal matrix, whose elements are the reciprocals of the square roots of the reliability coefficients used for correction (i.e. $1/\sqrt{r_{ii}}$, $i = 1, 2, \ldots, n$). As before, the rank and the number of factors remain unchanged. Thus, as Spearman noted, the vanishing of the tetrad-differences is unaffected by correction.

In what follows, since covariance has hitherto been so little used in psychology, I shall express my conclusions chiefly in terms of correlation.

CHAPTER XI

CORRELATIONS BETWEEN TESTS AND BETWEEN PERSONS

In almost every investigation of which factor-analysis forms a part, the initial table of data gives measurements for (*a*) a limited sample of the total universe[1] of persons in respect of (*b*) a limited sample of the total universe[1] of traits or tests. As we have seen in a previous chapter, we may base our factor-analysis on covariances (or correlations) either (*a*) between the several persons or (*b*) between the several tests. If our interest lies in a 'general factor' that may cover the universe of traits or tests (e.g. general intelligence, in the case of cognitive tests, general emotionality in the case of emotional traits), we shall naturally begin by correlating the tests or traits; if in a 'general factor' characteristic of the population of persons, we shall naturally begin by correlating persons (see above, pp. 175 f.): such factors are simply averages (or more accurately weighted sums) of the measurements for the traits and for the persons respectively. If our interest lies rather in the secondary factors, then, in theory at any rate, we have the option of either method of approach: in practice, provided the method of assessment permits it, economy of labour will generally suggest that we correlate whichever variable is overdetermined—traits if the persons are the more numerous, persons if the traits are the more numerous (see pp. 177, 260).

The relations between the two sets of factors can be exactly stated: if the averages for all tests and for all persons are the same, and if the method of analysis described above

[1] It is perhaps a little awkward to talk of a 'population' of tests or traits, though Fisher, Stephenson, and others use the term 'population' in that way. The logician's word 'universe' covers both cases quite naturally, and further implies that the tests or traits sampled should belong by implication to a definable class. Cf. [110], pp. 25, 332 *et seq.*, and p. 180 above.

is adopted, then, when duly normalized, *the factor loadings for persons obtained by covariating persons are identical with the factor measurements for persons obtained by covariating tests.* Similarly, the factor loadings for tests obtained by covariating tests are, when duly normalized, identical with the factor-measurements for tests obtained by covariating persons ([101], p. 82; for proof and worked examples, see Appendix II, Tables I-III). If the conditions specified do not hold, the relations between the secondary factors become a little more indirect, and vary according to the nature of the changes: usually, however, the correlation between the two sets of results is over ·9 ([114], p. 166, footnote 1). Accordingly, it still seems true to say that, in general, " except for minor differences of weighting, the non-general factors obtained by correlating persons are the same as those obtained by correlating tests or traits " ([137], p. 90). This conclusion is sometimes briefly referred to as the ' reciprocity principle.'

Here again my arguments have been questioned by Dr. Rhodes and Dr. Stephenson, and once more from opposite standpoints. Rhodes doubts the entire validity of applying the correlation calculus to persons as variables; but the results that he eventually reaches are almost identical with those obtained by my least-squares method of analysis from the corresponding correlations: hence, there can be little real difference between us (see his further *Memorandum* for the Examinations Inquiry Committee ([93], pp. 316–24)). Stephenson at first also expressed doubts about the method, but for different reasons: it seemed to imply correlating measurements in entirely disparate units, and would in any case produce (so he considered) innumerable common factors—a result quite incompatible with the two-factor theory to which he adhered (cf. above, pp. 177–8). But, on reading the draft of my *Memorandum* dealing with Rhodes' objections, he accepted the proposed mode of standardization and the adaptation of Spearman's theorems, and also provisionally agreed that the one method was " merely the obverse " of the other—i.e. (as he puts it) the two procedures are merely " complementary ways of analysing the same data " ([92], p. 304; [96], p. 195).

As a test of these and other issues (more particularly the reciprocity principle for abilities and temperamental tendencies) a programme of researches was drawn up, and carried out in our

laboratory with the help of several of our research students. Stephenson's interpretation of these results led him, as we shall see (chap. xix), to reject the 'reciprocity principle,' and to maintain that the 'type-analysis' by correlating persons could no longer be regarded as the 'obverse' of the older psychometric procedure, i.e. of the factorization of correlations between traits ('Postscript' to [96], p. 206). Indeed, so different in his opinion are the factors obtained by correlating persons and traits that he now considers "no analysis of a set of measurements can be considered complete until it has been analysed both ways" ([97], p. 361).[1]

With his criticisms of the principle I shall deal more fully in a later chapter. Meanwhile, as I have already observed (p. 191), in spite of minor changes in tests, traits, and statistical procedure, the various 'type-factors' obtained by Stephenson, whether cognitive, temperamental, or æsthetic, appear after all to be very similar to those already obtained with the alternative procedure. Let me add, I do not for one moment deny that it is possible to invent a factorial technique which will always lead to different sets of 'factors' when applied to correlations between tests and between persons respectively. But I see no advantage in multiplying such entities unnecessarily. On the contrary, the difficulty about the methods of factorial analysis hitherto proposed is that they allow us too much liberty in selecting our final factors. And one great advantage of the reciprocity principle, as it seems to me, is that, if we adhere to it on grounds of economy, it may enable us to narrow down that selection and even render it unique.

Since the foregoing paragraphs were written, Thomson's admirable book on *The Factorial Analysis of Human Ability* has appeared. In it he discusses, at some length and with his usual combination of lucidity, impartiality, and critical insight, what he terms the 'reciprocity of tests and persons' ([132], chaps. xiv and xix). He does not, as Stephenson appears to do, deny that the factorization of the two sets of correlations can *ever* lead to the same

[1] Among the results of the researches carried out under his guidance he cites those of J. I. Cohen. The data and methods then employed were hardly adequate to test the issue; and the completed investigations show a clear correspondence between the two sets of factors (a correlation of about ·9, according to his unpublished thesis, University of London Library). P. C. Hu has applied both procedures to Stephenson's own data, obtained with picture cards of Japanese vases; and obtains a similar agreement.

figures; but he is "of opinion that his (Burt's) is a very narrow case, and that the factors considered by Burt are not typical of those in actual use" (p. 213); and he ends his account of the method by affirming that "it would be wrong to conclude in general that loadings and factors are reciprocal in persons and tests," since "most of what we call association or resemblance between either tests or persons is due to something over and above the very special kind of residual association" shown in "Burt's doubly centred matrix" (pp. 218–9). With this latter argument as worded I fully agree: for "most of the association or resemblance (i.e. most of the correlation or covariance) between either tests or persons" is due to the first or dominant factor ('*the* general factor' for tests or for persons respectively). This by definition is the factor that accounts for the greatest amount of the total variance; and such factors are by hypothesis excluded from my 'reciprocity principle.' But, although he does not say so, the context seems to imply a belief that even the non-general (the secondary or bipolar) factors are not, as a rule, 'reciprocal' or equivalent. With that I cannot agree, unless by reciprocity we mean an exact reciprocity, regardless of the inexactitude which tests 'in actual use' always entail.

His chief criticisms of my argument may be summarized as follows. First, discussing my fictitious example ([101], pp. 90–4), he points out that "we could write down an infinity of possible raw matrices from which Burt's doubly centred matrix might have come" (p. 219). Thus, "to the *rows* of the matrix we can add any quantities we like, without altering the correlations between the tests, but making enormous changes in the correlations between persons." But, if that were done, we should be disturbing the relative difficulty of the tests: for in the rows the measurements are entered as deviations about the average for each test (or trait), and this implies that the tests are presumed to be of equal difficulty. This is a presumption common to nearly all current factorial work: (the only important exceptions are those cases in which the relative difficulty of the tests itself becomes an object of investigation, as, for example, in my investigation of the Binet-Simon scale, where the relative difficulty of the tests was studied by correlating persons). Again, he says, "we can add any quantities we like to the *columns* of the matrix of marks." But that would introduce differences in

'cleverness' among the persons tested, and so contradict my initial assumption, namely, that the "population is homogeneous as regards the 'general factor'": since, "if we desired to study 'specific' factors (i.e. type factors) more especially," it was natural to take "the means for the several persons as approximately equal from the outset" (*loc. cit.*, p. 72).

Secondly, he points out that perfect reciprocity can be demonstrated only if covariances, not correlations, are analysed. But, as he states elsewhere, "if a suitable set of units could be discovered, the practice of analysing covariances instead of correlations would have much to commend it" ([137], p. 76). The difficulties attending the use of covariances were freely granted; and I should never claim that in practice, despite our imperfect means of assessing intellectual abilities or temperamental traits, a mathematically *perfect* reciprocity can everywhere be demonstrated. When, however, the tests or traits are already in standard measure (as is the common custom), the covariances between them become identical with their correlations. The only trouble, therefore, arises over the correlations between persons. Here, I believe, unless special precautions are taken over the selection of the 'tests' (or what figure as such), correlating (as distinct from co-variating) persons may be hard to justify. But in any case the factors obtained from covariances will, with but little transformation, still serve to explain the correlations as well.[1]

Finally, in dealing with my last example [114], he criticizes more especially the units of measurement assumed. In this inquiry each trait was based on an assessment for about 20 elementary reactions; and the assessments for each reaction were in turn distributed by instruction in accordance with a normal curve having the standard deviation as unit (so far, at least, as the knowledge of the observer would permit). In theory this should lead to a variance for each composite trait roughly proportional to its complexity (i.e. a trait compounded by adding 20 *independent* reactions would show the widest amount of individual variation; a trait compounded of fewer independent or fewer observed reactions would show a smaller variation). Thomson agrees that "there is something to be said for the probability of real differences in variance." But, he adds, "it cannot be right to use a space whose metric is dependent upon accidental and irrelevant differences in the variable" (e.g. upon the lack of complete information in regard to

[1] Cf. [114], p. 179. Kelley appears to assume that the factors will be the same ([85], p. 1): but, as I have endeavoured to show in my review of his book, in the general case, this appears to be mistaken ([102], p. 193).

certain traits such as sex). I willingly acknowledge how difficult it is to obtain a satisfactory set of units, above all for emotional or temperamental characteristics. Thomson himself goes on to suggest the possibility of 'some system of natural units' (p. 293); and it will be of interest to see how this would work out in actual practice. At the same time, notwithstanding the admitted shortcomings in the rough methods of assessment that I myself adopted, I cannot believe that the 'accidental and irrelevant differences' in the actual variances have produced any great distortion in the essential nature of the factors. I am even tempted to suggest that a strict adherence to the principle of reciprocity might itself be made to indicate, or at least to confirm, what is the most convenient set of units.

Since the publication of his book, Thomson has generously withdrawn one point of criticism (cf. [136], p. 76); and, quite recently, has restated his own conclusions. At the close of the Reading Symposium on *Factor-analysis* he said: "Although I agree with a non-rigorous use of the reciprocity principle, I must emphasize that it only can be found rigorously in a very special sample of people who are all average in ability, and a very special sample of tests which are all of average difficulty. Prof. Burt will protest that it will hold approximately elsewhere; and I readily agree that it will do so if these conditions are not too flagrantly broken" ([136], p. 107). That expresses all that I want. I fully admit that a rigorously *exact* equivalence or 'reciprocity' can only be found under the conditions specified; it must be remembered, however, that other writers had previously denied that *any* equivalence could be found under *any* conditions.[1]

We shall return to this subject in Part III, where an illustrative example will be fully analysed and discussed.

[1] Superficially, no doubt, the reciprocity principle remains incompatible with Thomson's sampling theory, at least in its original form. Whereas other theories reduce the number of factors to a minimum, that theory implicitly assumes the maximum. Classifying the factors as suggested above (p. 104), Thomson's 'most probable factorial pattern' includes every possible kind of group-factor (cf. [132] fig. 12, p. 44): it invokes factors entering into any combination of 1, 2, ..., and n tests respectively. Thus the number of factors will be ${}^{n}C_1 + {}^{n}C_2 + \ldots + {}^{n}C_n = 2^n - 1$ in all. Now, when we apply the same 'analysis by overlap' to the correlations between persons, we obtain an entirely different number, viz. $2^v - 1$. Hence the factors, as thus defined, cannot be the same. However, as soon as we cease to identify each 'factor' with a fixed 'ability,' such incompatibilities no longer matter.

CHAPTER XII

GENERAL-FACTOR METHODS AND GROUP-FACTOR METHODS

(*a*) *Differences in Procedure.*—The earliest attempts at factor-analysis rarely ventured to establish more than a single general factor : such a factor, operating alone, would produce a correlation matrix of rank one—a ' hierarchy ' as it was termed—for which various criteria were proposed. Almost inevitably, so long as individual tests were used and only a dozen or so persons could be tested, the probable errors remained too high for the residuals to be significant when the general factor had been partialled out : hence no clear evidence could be found for more factors than one. But, as soon as the introduction of group-testing made it possible to diminish the sampling errors, it became manifest that the observed correlations could no longer be fitted to a strict hierarchical pattern, unless the tests were deliberately selected for this purpose, and that the discrepancies implied other factors besides the single ' general ' and the *n* ' specific.'

The first attempt at explicitly fitting a theoretical matrix to a set of observed correlations was, I think, that shown in Tables V and VI of my paper of 1909 ([16], pp. 161–2). The residual correlations, examined for further factors, were obtained in the way that has since become fairly general, namely, by simply subtracting a reconstructed hierarchy from the original correlations. The tests had been expressly chosen " to represent different types of mental process " ; and, as a result, " a small discernible tendency for subordinate groups of allied tests to correlate together " was noted. With increasing reduction of the probable error there seemed no reason why the process of removing one factor to reveal another should not be repeated. Thus, as I have pointed out above, it became natural to view the empirical table, not as an unsatisfactory approximation to a single hierarchy, but as the sum of a succession

of diminishing hierarchies, each representing a factor common to certain of the tests ([35], Tables XVIII–XXIII; [102], p. 189).

Now, according to the way the correlated variables were originally selected, these hierarchies, it was found, might either cover the *whole* table (as when the tests or persons correlated formed a random sample from a continuous distribution) or be confined to different *parts* (as when the tests or persons correlated formed a heterogeneous or a discontinuous set). Consequently, as we have already seen, the 'two-factor theory,' which merely singled out the one common factor contributing most to the total variance, developed on the one hand into a 'multiple-factor theory,' which envisaged several *general* factors, each common to all the tests or persons correlated, and on the other hand into what was sometimes called a 'three-factor theory,'[1] which looked rather for limited *group*-factors superposed upon a single general factor.

I shall call those methods which proceed by analysing the correlation matrix taken always as a whole 'general-factor methods,' and those which partition the correlation matrix into suitable sub-matrices, and then analyse these separately, 'group-factor methods.' (In previous publications they have sometimes been described as 'method *b*' and 'method *a*' respectively.) With the former, all the factors after the first have *negative as well as positive saturation coefficients*: with the latter, the factorial matrix has a large number of *zero*[2] *saturations* in each column, and for the rest contains only[2] *positive saturations* (cf. [93], p. 306).

Under the heading of general-factor methods may be placed those used by Spearman, Hotelling, Kelley, Thurstone (before rotating), and Stephenson (when analysing correlations between traits). Under the heading of group-factor methods may be placed Holzinger's most recent form of the bi-factor method, Thomson's alternative analyses of artificial correlation-tables obtained from dice or cards, Stephenson's usual procedure when analysing correlations between persons, and my own early efforts when demonstrating group-factors in educational abilities and in emotional traits.

[1] Cf. p. 139, above.

[2] At each point we ought to add 'within the limits indicated by the sampling error.' In actual practice a calculated saturation coefficient will never be *exactly* zero; and many of the non-significant figures may prove to be negative, where theory demands nothing but positive values. In what follows, however, to avoid confusing the statement of the main tendencies, I shall omit these obvious reservations.

When the tests or persons to be correlated are drawn from mutually exclusive classes, the group-factor describing one of the classes should show no overlap with the others. Several of Holzinger's earlier factor-patterns for tests and of Stephenson's factor-patterns for persons are of this type. But, with empirical data, if the calculations are carried far enough, the group-factors will, in point of fact, nearly always display some overlap. In most cases the first factor tends to spread over the whole sample of tests, so that (although for one or two tests its saturations may be strictly non-significant), nevertheless it is more convenient to describe it as a general factor. With educational tests, and usually with cognitive tests and with emotional assessments, I found group-factors overlapping like steps on a winding staircase, producing what I called 'cyclic overlap' ([30], [35]): granting that the 'central nervous system is integrated by a series of broadly distinguishable layers or levels, the higher levels serving to co-ordinate the lower,' such an overlap must be an almost inevitable result when a comprehensive sample of tests or traits has been selected. Thurstone's 'simple structure' assumes a free but irregular overlap of many group-factors, but no general factor whatever ([84], p. 151, [122], p. vii); actually, however, his tables are excellent examples of 'cyclic overlap' (if we rearrange his tests in the order: 5, 7, 3, 4, 9, 2, 6, 8, 1, his Table I, *loc. cit.*, p. 151, is a perfect instance). Thomson, who from the outset has emphasized the explanatory power of overlapping group-factors [34], has recently proved that the boundary conditions, indicating whether, in any particular case, general factors are equally indispensable or not, may be precisely defined [100].

The advocates of each of the two main methods have sharply criticized each other. Spearman, in a famous phrase, originally opposed the 'unifocal' or 'monarchic' doctrine of a single general factor to the 'multifocal' or 'oligarchic' doctrine that admits nothing but group-factors identified with mental faculties or mental types. As we have already seen, he does not deny all possibility of group-factors, but insists that, in contrast with the 4 or 5 general factors that have been isolated, such limited factors are both exceedingly narrow and exceedingly rare: in keeping with the two-factor theorem he prefers to call them 'overlapping specific factors.' Thurstone admits no 'general factors' whatever, even when found by his own procedure (except possibly as a residual): the factorial matrix must always be rotated before we can reach the primary abilities or traits. Holzinger, on the other hand, rejoins that "the methods employed by Hotelling and Thurstone give rise to com-

ponents or principal factors which are rarely factors in the sense defined" (by Spearman and himself)[1]: they are simply "artifactors"; "it is very seldom that we can establish their meaning" ([83], No. 5, p. 1).

But the results obtained by the two different methods appear incompatible only so long as we assume (with both parties to the controversy) that an abstract statistical factor, to have any intelligible meaning, must be identifiable with some concrete 'ability' or 'trait.' Suppose we are seeking a general expression for the physical differences between the sexes, and one investigator states his results by saying that the difference between their average heights is 6 inches, while another gives the men's height as + 1·5 S.D., and the women's — 1·4 S.D. The findings are not proved meaningless because the second declares that 6 inches is the height of no human being; while the former rejoins neither is — 1·4 S.D., because a negative height is impossible: both expressions are alternative but related formulations of one and the same fact. Nor is it difficult, I think, to demonstrate that a similar equivalence

[1] In the *Spearman and Holzinger Trait Studies* a factor is defined as follows: "The vanishing of the tetrads within limits of probable error furnishes the necessary and sufficient conditions for identifying a simple factor": ([83], p. 1); such a factor is to be conceived as a "unitary ability or trait." Unfortunately, in order that the tetrads shall vanish, it is necessary that the saturation coefficients for the group factors (or 'secondary factors,' as the writers prefer to call them) should be simple multiples of the saturation coefficients for the 'principal' or general factor (cf. [83], p. 2): but patterns constructed to conform with this condition would seem to be quite as artificial as any to be found in Hotelling or Thurstone (Thurstone himself, however, apparently regards it as at least a 'conceivable psychological situation' [84], p. 141). The unique collection of material accumulated and analysed by the several collaborators in the Spearman-Holzinger reports has since led to successive and important modifications of the original theory; and the recent account of the 'bi-factor method' in its latest form, given by Holzinger in his remarkably clear and suggestive manual [106], brings it much closer to the methods of other workers. The scheme for the generalized factor-pattern now seems identical with that accepted in the report of the Examinations Inquiry Committee, and the terminology more nearly in line with that prevalent in this country. Thus, the classification and sub-classification of factors given in the new *Students' Manual* ([106], p. 12) appear identical with the 'four-factor theory' previously given in my own *Memorandum* and elsewhere ([93], p. 259); and the factor-pattern in Holzinger's latest article ([107], p. 43) is identical with that in my own table ([93], p. 264, Table 135). Here, therefore, once again independent workers, starting from very different standpoints, seem gradually to have arrived at much the same general result.

holds good of measurements in terms of 'group-factors' and of 'general factors' respectively.

(*b*) *Relations between the Saturation Coefficients.*—Correlation tables that lend themselves most obviously to the 'group-factor' mode of analysis usually arise when the tests that have been correlated (or the persons, if we begin by correlating persons) form a discontinuous selection: (e.g. tests representing different mental levels or functions [16], [82], examinations in different school subjects [35], persons of different types of imagery [116], etc.). In such cases the coefficients can be reassorted so that exceptionally high correlations that break the hierarchy cluster in blocks along the principal diagonal, while the remaining rectangles fit more appropriately into the general hierarchical order (cf. [30], Table I; [35], Tables XVIII, XX, XXII, XXIII). Thus arranged, the whole array of coefficients can be partitioned into square and oblong submatrices, grouping together those tests or traits which belong to much the same categories and are therefore attributable to the same 'group-factor.'

To illustrate the resulting scheme at its simplest, let us suppose that both the variances for the different factors and the saturation coefficients for the different tests are everywhere equal (say ·5 throughout). With 4 group-factors each found in two tests, we should then obtain an 8×8 correlation table such as the following (Table II), which may be partitioned as shown and regarded as a 'compound matrix' ([26], p. xii; [73], pp. 5–8). The rank of the matrix is evidently only 4; but the 4 conspicuous clusters, and the positive correlations outside the clusters, suggest an analysis into 5 factors. The non-mathematical student will most easily grasp the relations between the alternative modes of factorization if we now apply them to a simple arithmetical example, such as this.

(*a*) To determine the factors by the group-factor method we may follow the procedure I have outlined elsewhere ('method *a*' [93], p. 306; [116], p. 339).

On applying this procedure to an empirical table, the first step, as we have seen, is to group the correlations so that the enlarged

TABLE II

INTERCORRELATIONS BETWEEN FICTITIOUS TESTS

Test	1	2	3	4	5	6	7	8
1	·50	·50	·25	·25	·25	·25	·25	·25
2	·50	·50	·25	·25	·25	·25	·25	·25
3	·25	·25	·50	·50	·25	·25	·25	·25
4	·25	·25	·50	·50	·25	·25	·25	·25
5	·25	·25	·25	·25	·50	·50	·25	·25
6	·25	·25	·25	·25	·50	·50	·25	·25
7	·25	·25	·25	·25	·25	·25	·50	·50
8	·25	·25	·25	·25	·25	·25	·50	·50

coefficients form square blocks along the leading diagonal, as in Table II. If the whole table has been arranged as nearly as possible in a hierarchical order, i.e. according to the totals of each column, the further readjustments can usually be made by simple inspection —mentally comparing the contours of each successive row: if there is any doubt or difficulty, these contours may be actually plotted and sorted according to their resemblances; and finally, for the few doubtful cases, the inter-row or inter-columnar correlations can be explicitly computed: thus, in Table II, the correlation between rows 1 and 2 is obviously unity and between rows 2 and 3 negative ([116], p. 346).[1]

Here, to simplify the arithmetic still further, let us pool the tests in each group (i.e. take the sum of the figures in each rectangle). This reduces the matrix to an order equivalent to its rank, i.e. 4×4, and incidentally gives easy figures (2 and 1) for our schematic illustration. The totals in the leading diagonal, being each of them enlarged by its own group-factor, will now be omitted, just as we omit the magnified reliability coefficients from an ordinary hierarchical table; and the variances for the first factor (estimated by simple smoothing or some equivalent device) will be inserted in their place.

[1] Plotting the contours is in some instances more reliable as well as more speedy, for the regression of one row on another may be non-linear. As I use it, the 'inter-columnar criterion' is calculated *without* adjusting for the means of the correlated rows: this implies that the correlation matrix is itself 'squared,' just as the measurement matrix was 'squared' to obtain the correlations. The resulting 'square' is also required for the general-factor method as calculated by higher moments.

The saturation coefficients can then be calculated either by the summation method or by that of least squares: in this instance both methods lead to the same figures, given for the pooled tests in the first column of Table IIIA.

If the reader will partial out the first factor in the usual way by subtracting the expected correlations from the observed, the residuals will be found to yield a symmetric 'compartite matrix,'[1] having positive figures in the squares along the leading diagonal and zero elsewhere. Each of the square sub-matrices (technically termed 'parts') can now be treated as a separate hierarchy, and factorized. If the entire matrix has 4 hierarchical 'parts,' its rank will be 4 ([26], p. 7); the completion of the analysis will therefore yield 4 'group-factors' in addition to the first 'general factor.'

To save space, instead of tabulating all the values for the separate tests, I print tables of the pooled values only.[2] The complete factorial matrix as obtained by this method will be the 4×5 matrix formed by the five columns of Table IIIA (F_a). The whole has the appearance of a vertical column-matrix to which a diagonal matrix has been subjoined, each element representing a sub-matrix of one or more short columns.

Matrices having the form shown in Table IIIA are by no means unusual in the resolution of compartite matrices and in the theory of canonical reductions generally (cf. [26], II, p. 12 f.). By the psychologist, however, both the compartite scheme and the pattern into which it is resolved will be recognized as repeatedly occurring in analyses of correlation-tables where the investigator has first removed a general factor and then extracted a series of positive non-overlapping group-factors. Examples are to be found in my early analyses of emotional traits and educational abilities ([30], p. 695; [33], p. 37; [35], Tables XX, XXIII), and later in Alexander's analyses of correlations between intelligence tests of various types ([82], Tables on p. 148 and XXVI–XXIX), and most recently of all

[1] A 'compartite matrix,' when rearranged in standard form, is a compound matrix all of whose sub-matrices vanish except those lying along the leading diagonal. These non-vanishing sub-matrices are called its 'parts.' In general, they will not necessarily be square; but they must be square if the matrix is symmetrical ([26], II, p. 7).

[2] Incidentally this will serve to illustrate the treatment of 'specific factors' by the general-factor method. To obtain the saturation coefficients for the separate 'tests' the figures in the factor matrices must, of course, here be halved throughout. For a more detailed example, see pp. 477 f.

in Holzinger's 'hollow staircase' patterns (e.g. [83], No. 6, Tables 9–13). Davies and I found similar patterns when correlating heterogeneous groups of persons selected to represent well-marked mental types. As in Thurstone's 'simple structure pattern' ([84], p. 151), the final factor-pattern contains *no negative saturations* and a *maximum of zero saturations*. But this older scheme departs from Thurstone's requirements in two respects: first, the number of factors (here 5) exceeds the rank (here only 4); secondly, one of the factors—the first—has no zero saturations at all. We have thus what I have called the 'prop-ladder pattern' as distinct from the simple 'step-ladder pattern' of Thurstone.[1]

(*b*) For comparison let us analyse afresh the same correlation table by the 'general-factor method' ('method *b*'). We now retain the diagonal values from the start, with the natural result (as we shall see) that the effect of each 'group-factor' is distributed over the first as well as over the later factors, and over a much larger number of tests.

In the present instance we could take the pooled values and form a 'characteristic equation' ([101], p. 78), thus reaching the required results not approximately but exactly and directly; alternatively, if we apply either the summation method (which is here identical with a Thurstone analysis) or the method of least squares with successive approximation (which is here identical with a Hotelling analysis), we still arrive in the end at precisely the same figures (Table IIIB, F_b). There are now 4 factors instead of 5: the factor with the highest variance has positive loadings for every test; the remaining three are bipolar, i.e. they show *negative* as well as positive saturations.

Now if our test-assessments were measures of temperamental traits, we should feel no hesitation over assigning an immediate psychological interpretation to the general factors thus obtained: factor i might be identified with 'general emotionality'; factor ii with a bipolar tendency, making for 'extraversion' when positive and 'introversion' when negative; and so on (cf. [114], pp. 182–3). If, however, the tests were measuring cognitive traits, then most psychologists would apparently be unwilling to interpret the bipolar factors as they stand. Factor i could perhaps be identified with 'general intelligence': but what meaning, it has been asked, can

[1] Alexander's analyses, though following Thurstone's model, usually have a general factor prefixed, and so yield a 'prop-ladder' pattern. Thurstone's group-factors usually show considerable overlapping, which does the work of the general factor.

possibly be assigned to 'a factor which improves certain test-performances when it is not merely absent, but actually negative' or to 'an ability whose possession is a detriment to performance'? (Cf. [82], pp. 99, [84], pp. 161, 165–6, [122], p. 71). The procedure recommended,[1] therefore, would doubtless be to 'rotate the axes' until a simpler and more intelligible factor-pattern was secured. For this purpose the plan usually followed[2] is to plot graphs for each pair of factors, fit fresh axes at right angles to one another by eye, and measure the angles between them and the original axes with a protractor. But so rough a device must obviously yield somewhat arbitrary and inexact results.

The relation between the results of the group-factor method and those of the general-factor method can be expressed by a simple transformation matrix. Elsewhere I have described how such a matrix may be computed, and have shown how it yields a direct procedure—far more exact than the usual graphical devices—for rotating axes to abolish the negative saturations [116]. As calculated for the present figures, it is shown in Table IIID (T). It is evidently a 5×5 orthogonal matrix with one of its rows deleted: hence the transformation matrix for the inverse operation (which we should denote by T^{-1}, were T non-singular) is simply the transpose of the previous, namely T' (Table IIIc).

In order to illustrate the derivation and structure of rotation matrices such as T, let us suppose that (with Thurstone and Alexander) we have started by analysing our correlation table according to the 'general-factor method,' and now desire to abolish the negative saturations and to maximize the zero saturations in the factorial matrix so obtained, namely F_b. In theory the problem is familiar enough: it is evidently that of obtaining for F_b, by a series of elementary transformations, an 'equivalent matrix,' fulfilling the specified requirements. The simplest procedure is to keep operating on the columns (or rows) of the given matrix (and on the results of our operations) by the ordinary methods used in the reduction of determinants, until we reach a pattern of the type desired. If the

[1] Employed or suggested by Thurstone, Thomson, Alexander, Guilford, Stephenson.

[2] "The graphical method of rotating in one plane at a time is still probably the best. . . . But the graphical method is not ideal" (Thurstone, 1938, [122], pp. 72–3).

proper conditions are observed, the procedure will of itself ensure that the resultant transformation is orthogonal. The above table lends itself to a simple illustration showing how the peculiarities of the transformation matrix arise.

(i) We begin with the bottom row of F_b, since this contains the largest number of negative saturations. These can all be reduced to zero by subtracting suitable submultiples of the first column (which contains positive figures throughout) from all the other columns in turn. This means post-multiplying F_b by a matrix of the type :

$$\begin{bmatrix} 1 & -\frac{f_{12}}{f_{41}} & -\frac{f_{43}}{f_{41}} & \dots \\ 0 & 1 & 0 & \dots \\ 0 & 0 & 1 & \dots \\ \dots & \dots & \dots & 1 \end{bmatrix} = \begin{bmatrix} 1 & \frac{1}{\sqrt{15}} & \frac{\sqrt{2}}{\sqrt{15}} & \frac{\sqrt{6}}{\sqrt{15}} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Since $f_{42}, f_{43} \dots$ are negative, this produces a row of positive figures in the top line.

(ii) But this first operation inevitably raises the zero saturations in the top right-hand triangle of the initial matrix to positive values. If we are to maximize the zeros, these must be abolished. We proceed as before ; and now the operation will be equivalent to multiplying by a matrix including a corresponding triangle of negative multipliers, viz. :

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & -\frac{1}{2\sqrt{2}} & -\frac{1}{\sqrt{6}} \\ 0 & 0 & 1 & -\frac{1}{\sqrt{3}} \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

(iii) When all the negative saturations and all the eliminable positive saturations have been annulled in this way, we multiply by a diagonal matrix :

$$\begin{bmatrix} \frac{2}{\sqrt{5}} & 0 & 0 & 0 \\ 0 & \frac{3}{2\sqrt{3}} & 0 & 0 \\ 0 & 0 & \frac{2}{\sqrt{6}} & 0 \\ 0 & 0 & 0 & \frac{1}{\sqrt{2}} \end{bmatrix}$$

(and in the present instance add a supplementary column) to keep the total variance of the rotated matrix the same as that of the unrotated matrix.

Combining these successive post-multipliers into a single post-factor, we obtain the transformation matrix (T) shown in Table IIID. The reason for its peculiar structure will now be evident, if we recall the way in which the initial factor matrix F_b is obtained. The first column of positive figures was the result of averaging a positive correlation matrix; the remaining bipolar columns represent weighted deviations about that average. Since each later column is uncorrelated with the preceding, the negative section of the preceding column is always followed by a bipolar section in the later column[1]: this leads (with a suitable interchange of rows) to a triangular arrangement of the negative saturations. Moreover, since we extract the dominant factors first, the variance of each column is progressively reduced. To abolish the triangle of negative saturations will require a triangular (or truncated triangular) transformation matrix; and the product of the transformation will itself be triangular, and (since the interior of the triangle consists of zeros produced by the endeavour to maximize them) will exhibit a step-ladder pattern.[2]

On applying the transformation T to our second factorial matrix (i.e. post-multiplying Table IIIB by Table IIID), we obtain the product matrix set out in Table IIIE. It will be seen at once that the product is identical with the first factorial matrix (Table IIIA), which was obtained directly by 'method *a*'—the 'group-factor' or sub-matrix method. Thus (as I have elsewhere tried to show by applying the same procedure to the tables of Alexander and other writers) *it is in fact unnecessary first to factorize the correlation matrix by a general-factor method and then*

[1] In actual work the third column may be doubly bipolar, the fourth quadruply bipolar, and so on. But I am here dealing with the main scheme in its simplest form. It will be sufficient to point out that the additional bipolar sections always contain low figures, and so do not disturb the predominating pattern.

[2] A similar result may be obtained by operating either directly on the correlation matrix, or, more laboriously but more directly still, on the initial matrix of measurements. The use of triangular matrices in the reduction of given matrices to orthogonal and canonical forms is well known. Cf. Schmidt, *Math. Annal.*, LXIII (1907), pp. 442 *et seq.*; Jacobi, *J. f. Math.*, LIII (1857), pp. 265 *et seq.*, and [73], pp. 64, 96, 105 *et seq.*

progressively rotate the results in order to get group-factor saturations : we can reach the latter in a single step.

TABLE III

RELATIONS BETWEEN FACTORIAL MATRICES OBTAINED BY GROUP- AND GENERAL-FACTOR METHODS RESPECTIVELY

						TABLE IIIc Hypothetical Transformation Matrix (T')				TABLE IIId Calculated Transformation Matrix (T)				
						$\frac{2}{\sqrt{5}}$	0	0	0	—	—	—	—	—
						$\frac{1}{2\sqrt{5}}$	$\frac{3}{2\sqrt{3}}$	0	0	$\frac{2}{\sqrt{5}}$	$\frac{1}{2\sqrt{5}}$	$\frac{1}{2\sqrt{5}}$	$\frac{1}{2\sqrt{5}}$	$\frac{1}{2\sqrt{5}}$
						$\frac{1}{2\sqrt{5}}$	$-\frac{1}{2\sqrt{3}}$	$\frac{2}{\sqrt{6}}$	0	0	$\frac{3}{2\sqrt{3}}$	$-\frac{1}{2\sqrt{3}}$	$-\frac{1}{2\sqrt{3}}$	$-\frac{1}{2\sqrt{3}}$
						$\frac{1}{2\sqrt{5}}$	$-\frac{1}{2\sqrt{3}}$	$-\frac{1}{\sqrt{6}}$	$\frac{1}{\sqrt{2}}$	0	0	$\frac{2}{\sqrt{6}}$	$-\frac{1}{\sqrt{6}}$	$-\frac{1}{\sqrt{6}}$
						$\frac{1}{2\sqrt{5}}$	$-\frac{1}{2\sqrt{3}}$	$-\frac{1}{\sqrt{6}}$	$-\frac{1}{\sqrt{2}}$	0	0	0	$\frac{1}{\sqrt{2}}$	$-\frac{1}{\sqrt{2}}$
Tests														
1 & 2	1	1	0	0	0	$\frac{\sqrt{5}}{2}$	$\frac{3}{2\sqrt{3}}$	0	0	1	1	0	0	0
3 & 4	1	0	1	0	0	$\frac{\sqrt{5}}{2}$	$-\frac{1}{2\sqrt{3}}$	$\frac{2}{\sqrt{6}}$	0	1	0	1	0	0
5 & 6	1	0	0	1	0	$\frac{\sqrt{5}}{2}$	$-\frac{1}{2\sqrt{3}}$	$-\frac{1}{\sqrt{6}}$	$\frac{1}{\sqrt{2}}$	1	0	0	1	0
7 & 8	1	0	0	0	1	$\frac{\sqrt{5}}{2}$	$-\frac{1}{2\sqrt{3}}$	$-\frac{1}{\sqrt{6}}$	$-\frac{1}{\sqrt{2}}$	1	0	0	0	1
Factors	i	ii	iii	iv	v	i	ii	iii	iv	i	ii	iii	iv	v
	TABLE IIIa Group-factor Matrix (F_a)					TABLE IIIb General-factor Matrix (F_b)				TABLE IIIe Factor Matrix obtained by Rotation ($F_bT = F_a$)				

Matrices approximating to the curious triangular form illustrated by Tables IIIB and IIIc are by no means uncommon in factor-analysis. I have already ([93], p. 307) drawn attention to their frequent appearance, and have indicated how they are related, as regards algebraic origin, to the triangular matrices that appear in the canonical reductions described in nearly every textbook, and as regards logical significance to the triangular tabulations that result from repeated classification by dichotomy. The principle, I venture to think, might be more directly exploited for practical calculation as well as theoretical interpretation.[1]

(*c*) *Relations between the Factor-measurements.*—Nor is it true that the factors originally reached by a general-factor method of analysis are necessarily devoid of meaning, or that, ' no matter how the correlational matrix is factored, the axes must always be rotated before any psychological interpretation can be made ' ([16], p. 74). If we grant a psychological meaning to the simplified factor-pattern obtained *after* transformation, then (as I think will now be

[1] Let me give an instance in which much the same principle may be applied to a situation not infrequently arising in actual practice. Suppose, for example, one of our tests is an almost perfect test of $g = s_1$ (say). By the repeated application of an obvious modification of the formula for partial correlation :

$$r_{ks_j} = \frac{r_{kj} - \sum_i r_{js_i} r_{ks_i}}{\sqrt{1 - \sum r^2_{js_i}}}, \qquad i = 1, 2, \ldots, (j-1)$$

we can eliminate first g and then $(n-1)$ other factors in turn thus obtaining saturation coefficients for n orthogonal factors : these saturation coefficients will remain unchanged, even if further tests are added to the battery. Where, as so often happens, our tests show a cumulative complexity, this form of analysis may well be employed, though (for reasons given in the paragraph cited) it seems unsuited for correlations between persons. The procedure is equivalent to a Lagrange transformation—the ' rational reduction of a quadratic form ' to a sum of squares by a succession of non-singular linear ' transformations ' (for proof and arithmetical illustration, see [15], p. 132 ; cf. [73], pp. 85–6). The cumulative transformation matrix is then triangular; and is admirably adapted to express partial or cyclic overlap. If, on the other hand, the specific factors do not overlap at all, so that there are no group-factors, the resulting zeros convert the triangular matrix into the hollow factorial matrix (a simplified ' prop ladder') which, on removal of the perfect test, represents the Spearman two-factor theorem. Thus we may regard the solid triangular matrices as produced by a filling-in of the Spearman factorial matrix to allow for overlapping.

manifest) the transformation itself is of such a nature that the initial factor-pattern obtained *before* transformation must also bear a related interpretation. This can be confirmed by considering the relations between the two sets of factor-measurements.

Let us suppose that we have reached, directly by the group-factor method or through an appropriate transformation, a simple and satisfactory factor-pattern. The five factors themselves we shall endeavour to identify from the tests or traits we have been correlating; let us call them (to keep the illustration concrete) (i) general intelligence, (ii) verbal, (iii) manual, (iv) arithmetical, and (v) artistic capacity, respectively. Let us further suppose that we have calculated the hypothetical measurements of our tested population for these five group-factors,[1] and desire to deduce their measurements for the equivalent general factors. Which of the group-factors must we take, and in what proportions must we combine them, in order to obtain estimates in terms of a given general factor? Call the first set of factor measurements P_a (i.e. 'population matrix for group-factors as obtained by method *a*') and the second P_b (i.e. 'population matrix for general factors as obtained by method *b*'). We require a transformation matrix T, such that $TP_a = P_b$. Curiously enough, T as thus defined turns out to be identical with the rotation matrix T already obtained.[2] In other words, T not only converts F_b into a 'primary' structure F_a; it also tells us how to weight and add or subtract the group-factor measurements in P_a so as to obtain the general-factor measurements in P_b.

It thus becomes evident that the rotation matrix already obtained not only enables us to transform the one set of saturation coefficients into the other, *it also describes the relation between the first set of factor-measurements and the*

[1] The first of the five would ordinarily be called a 'general factor,' or rather '*the* general factor'; but, to save circumlocution, I may perhaps be permitted to use the phrase 'group-factors' to mean factors obtained by the 'group-factor method of analysis.'

[2] I have given a formal proof elsewhere ([116], pp. 339–75); but the following perhaps makes the relations somewhat clearer. Adopting the same notation as before and using the weighting equation xxxvii ([93], p. 299), we have:

$$P_a = F'_a R^{-1} M = (F_b T)' R^{-1} M = T' F'_b R^{-1} M = T' P_b.$$

But $TT' = I$. Hence $P_b = TP_a$.

second: and this relation it demonstrates to be essentially a relation between a set of positive group-factors, on the one hand, and a corresponding set of bipolar factors, on the other, obtained by taking weighted differences between the former.

Our concrete example will make this clearer. To find p_{i1}'s measurement in the *new* first factor, T instructs us to take a large positive amount of the *old* first factor ('general intelligence') and add small positive amounts of all the others. To find his measurements for the new second factor, we take none of the old first ('general intelligence') because the new factor is highly specific; but we take a large positive amount of the old second factor (estimated from the verbal tests of reading and spelling) and smaller negative amounts of the other three factors (which were based on the non-verbal tests). In the case we are imagining, therefore, the new second factor will depend on the *difference* between the man's measurement in the verbal tests, on the one hand, and his measurements in the non-verbal tests, on the other. Similarly with the other factors. The bipolar character of all of them is thus at once explained: their saturation coefficients will now range, not from 0 to + 1, but from + 1 (denoting, e.g., complete verbality) to − 1 (denoting complete non-verbality), and, between these limits, will express the difference between the two contrasted tendencies.

Although bipolar factors are generally supposed to be devoid of intelligible meaning when they are obtained by correlating cognitive *tests*, nevertheless, if we were to correlate precisely the same set of measurements by *persons* instead of by tests, we should accept such bipolar factors as quite natural: we continually class children into 'verbal' and 'non-verbal' types, 'practical' and 'non-practical' types, 'mathematical' and 'non-mathematical' types, and so on, and such antithetical pairs logically imply a bipolar principle. Now I have endeavoured to show that, whether our analysis starts by correlating persons or by correlating traits, the same set of factors will be discernible in the same set of measurements. It follows that the bipolar factors must be equally intelligible when we start with the tests instead of with the persons tested. As I have argued elsewhere ([128], p. 69), if we grant that factors are essentially principles of classification, then *bipolar factors denote the corresponding principles of 'dichotomous' classification, while the corresponding group-factors deducible from them denote the corresponding principles of 'manifold' or co-ordinate classification.* And no one would contend that the dichotomous classification of, say, 'Porphyry's tree' was necessarily devoid of meaning, because half the subaltern gener a

were negative, and must first be converted into positive or null classes.

The transformation matrix that I have described provides the link between the results of factor-analyses by the general-factor method and the group-factor method respectively. The matrix is triangular. In practice it tends always to show the following features—each plainly exemplified by the ideal scheme of Table IIID : (i) diminishing positive figures in the top row ; (ii) diminishing positive figures in the diagonal ; (iii) zeros below the leading diagonal ; (iv) small but increasing negative figures above the leading diagonal (except for the top row) ; and—though this is less uniformly present—(v) figures in all rows except the first adding up to zero. Its general effect will be evident from a consideration of these characteristics. And if we study the process of post-multiplying F_a (Table IIIA) by the inverse, namely T' (Table IIIc), we shall at once understand how the several ' general factors ' might have been derived from the ' group-factors,' instead of *vice versa*. Each column of T' in turn (excluding the first) takes one of the group-factors, and gives to this factor a high positive weight, and to those factors that succeed it low negative weights, the total of which will approximately balance the positive. The result is that the saturation coefficients of these succeeding factors are subtracted from those of the preceding factor. It follows that, when applied to a population matrix obtained by the group-factor method, this inverse transformation must in effect substitute for the positive group-factors a series of bipolar difference-factors. Hence *the real purpose of the subsequent rotation* (upon which Thurstone insists) *is to convert these difference-factors back into the positive factors from which they were implicitly derived.*

The factorial and transformation matrices on which my argument has here been based are admittedly schematic. But, once the types have been recognized in their simplicity, matrices showing the essential characteristics of F_a, F_b, and T will be easily discerned in published investigations ; and, by the use of matrix notation, the arithmetical example

I have taken can be readily expressed in algebraic form and so generalized.

Illustrative Application.—Perhaps the best way to make the argument clear and plausible to the general reader will be to apply it to a concrete case. Let us take the example used by Thurstone in *Vectors of the Mind* [84] to illustrate the need for rotation and the methods to be used.[1] He chooses a correlation table obtained by Brigham for fifteen cognitive tests of a familiar type. Analysed by the centroid method the table yields three significant factors—one general factor and two bipolar (Table 2, p. 167). Thurstone then argues that " it is an error frequently made to attempt a psychological interpretation of the factors in [such a] table : . . . the orthogonal reference axes obtained by the centroid method must be rotated into a new set of reference axes before any psychological interpretation can be made."

Accordingly, after describing what in his view is needed for a ' simple structure ' such as could be interpreted in psychological terms, he shows how its requirements can be secured by a ' graphical method of rotation.' The figures thus obtained form a new factorial matrix (Table 4, p. 169). This also shows three factors. But now for each factor only one-third of the saturation coefficients (or thereabouts) are positive and significant. These are all above ·22 ; the remainder, including the negative coefficients, are (with few exceptions) all below ·12. With the factors in this final form we are at last in a position to " consider tentatively the psychological nature " of the ' primary abilities ' underlying the 15 tests : the three factors, it is concluded, represent respectively verbal ability, numerical ability, and a visuo-kinæsthetic ability, which is perhaps not primary but complex.

If, however, my views are right, this tripartite classification could have been inferred quite as well—indeed, even better—from the original factorial matrix obtained by

[1] Pp. 108–19. Five factors were actually extracted : but the factor-loadings for the fourth factor " are not large enough to justify serious consideration " (p. 167) and " the contribution of the fifth factor can be ignored " (p. 113).

simple summation (Table 2, p. 167). In that matrix the first factor has positive saturations throughout, all over ·40. To determine the psychological nature of this and other factors, I should rely, not on 'inspection of the tests,' but on introspection by the testees.[1] A long experience with similar tests, in which introspective reports have been obtained both from children and from trained students, convinces me that processes of relational perception are regularly used for the successful solution of every one of the tests in question here. I am not, however, concerned to defend the existence of a general cognitive factor, but merely to note (i) that we have no right to take its non-existence for granted, and (ii) that its adoption would remove the anomalous saturation coefficients that still remain in Thurstone's rotated factorial matrix, and would thus achieve a much more complete approach to 'simple structure' for the three more specialized factors that he seeks to establish.

The remaining factors in the unrotated matrix are bipolar. The second has six positive saturation coefficients and nine negative. If the distribution of these coefficients is plotted on a graph, a remarkable discontinuity appears: there is a wide interval separating the positive coefficients from the negative, and almost as wide a gap separating the

[1] A grave defect of nearly all the studies by factor-analysis during recent years is the complete neglect of introspection. Thus, Thurstone's general procedure ([122], p. v) is to select tests from certain tentative categories representing certain postulated abilities: factorial analysis is then used to confirm the hypotheses on which this selection was based—the hypotheses, namely, that tests 8, 29, 37, etc., are solved by "finding a rule or principle," tests 8, 25, 40, etc., by deduction, tests 6, 45, and 55, etc., by visualization and so on: but these are points to be determined primarily by introspection. For the same reason it seems to me unwise to assume (as Stephenson apparently does) that by simply tabulating tests according to their 'material' (or 'content') and their 'form,' without any corroborative introspection, we can reach a classification of the processes they elicit. Anyone who has systematically discussed his tests with the children he has tested will realize that a problem which is regularly solved by children of nine by one method may be solved at the age of twelve by an entirely different procedure. I myself was once assured by a psychological investigator that I had a 'good visuo-spatial imagery' because I excelled most other testees in tasks designed to test that ability: actually, I solved them all by rapidly converting them into verbal form—a trick which I long ago acquired to compensate for my weak visualization.

first three tests from the next three. On turning to their description, we find that the first six tests are all verbal, whereas the content of the remainder is essentially non-verbal. This factor therefore represents a dichotomous classification of the fifteen tests into two main categories—verbal and non-verbal.

The third factor is, as usual, ' doubly bipolar ' : that is to say, its saturation coefficients for the first six tests (which all had *positive* saturations with the second factor) are half negative and half positive, and so add up approximately to zero ; and again the saturation coefficients for the remaining nine tests (which all had *negative* saturations with the second factor) are four of them positive and five of them negative, and so once again add up approximately to zero.[1] This third factor, therefore, really indicates two distinct dichotomous sub-classifications : (i) first of all, it divides the nine non-verbal tests into (*a*) a sub-group of four, which turn out to be all numerical, and (*b*) a sub-group of five, which turn out to be all of the performance type, i.e. visuo-kinæsthetic ; (ii) and secondly it also divides the six verbal tests into two antithetical kinds, and so explains the second gap already noted. *But the principle of division is no longer the same.* A glance at the nature of the tests reveals the true reason. The first three tests, which are here given negative saturation coefficients, are all labelled ' Opposites tests ' ; the remaining three are verbal intelligence tests of a more miscellaneous character (Analogies, Definitions, etc.). It would be absurd to argue that the Opposites tests are all visuo-kinæsthetic and the remainder numerical.

Now let us compare this mode of classification with that disclosed by the new factorial matrix reached after rotation. We see that the new classification embodied in the ' simple structure ' agrees precisely in regard to two of the lines of subdivision ; but the last of them has disappeared, or at

[1] The totals are not exactly zero, but about ·03. On turning back, however, to the mode of computation, we discover that the discrepancy is evidently due to the insertion in the leading diagonal of the highest correlation instead of the figure demanded by the general pattern (see below, Appendix I, p. 450). We shall find in a moment that this has apparently influenced the interpretation offered.

any rate become masked. In fact, except for one small point, the new factorial matrix is exactly similar to those obtained by some of us with what was termed the ' group-factor method ' : it shows a step-like set of saturations indicating three sharply distinguishable group-factors. There is, however, a strange overlapping between two of them, namely, between Thurstone's numerical and verbal factors. His explanation is that the second set of verbal tests (what I have called the miscellaneous sub-group) show " some of the precision and restrictiveness of numerical work " ; and he goes on to suggest that " the number factor," which he supposes to be common to these three verbal tests as well as to the four numerical tests, may after all not be concerned with number as such, but with " some kind of facility for logical or other restrictive thinking of which numerical work is only a good example." That, however, comes very near to expanding the numerical factor into a general factor of relational thinking. This third factor, indeed, has positive coefficients for nearly all the tests : so that the attempt to dispense with a relatively general factor is, after all, by no means complete or convincing.

I venture to suggest that we have here an instance of the way in which the assumptions prevalent among most factorists tend to warp their psychological interpretations. These assumptions are, as we have already seen, that a single factor must represent a single ability and that it must therefore be interpreted in the light of its positive saturations only. Such a view was natural enough in the days when, owing to high probable errors, but one or two factors at most could be extracted : for the first factor nearly always showed nothing but positive saturations and could at once be explained as a general ability ; and the second factor, which, as I should maintain, really denotes a contrast between *two* abilities, could be identified with whichever of the two had positive signs. This mode of interpretation, however, is now extended to the third and fourth factors that can often be distinguished with the more reliable data of the present day. But in these further factorial columns we are dealing with factors which draw a contrast, not between one pair of abilities (or, as I should prefer to say, one pair of test-groups) but between two or more pairs. If, following Thurstone's principles, we eliminate, reduce, or otherwise ignore the negative saturations, then

we may be left with two sets of positive saturations which really belong to two different test-groups : the principle of one factor, one ability, however, leads the investigator illogically to merge these two groups into one.

But there is a further fallacy. Since all summation-factors except the first express merely a contrast, the choice of signs is arbitrary. For example, in Thurstone's table the last nine tests fall into two sharply distinguishable groups, namely, the four 'numerical tests' and the five 'performance tests.' Thurstone's third centroid factor, with its opposite signs, brings out the difference quite plainly. If, however, our positive signs are to indicate, not an abstract distinction but a concrete ability, how are we to proceed? Are we to allot the positive sign to the first four tests on the ground that their coefficients are somewhat larger, and regard the factor as essentially a 'numerical' factor? Or are we to allot the positive sign to the last five tests on the ground that they form the majority, and so regard the factor as essentially a 'visuo-kinæsthetic' or 'performance' factor? Thurstone himself, it would seem, seeks to combine both principles by giving the positive sign to that group of tests which yields the largest *total* saturation, whether due to number or to size. But the slight divergence in the two totals is, as we have noted, due simply to the rough estimate of the communalities, and the continual substitution of the largest coefficient in the column ; with an exact estimate the totals would be equal.

Yet, arbitrary though it is, once the choice for the positive sign has been made, that seems to determine the subsequent interpretation.[1] If the saturations which are arbitrarily made negative are small, they are likely to be ignored. If they are large, the rotation will perhaps not entirely suppress them, but postpone them to a later factor. A glance at the results, however, shows that with the postponement they are very apt to lose their importance. Here, for instance, Thurstone has chosen to give *negative* signs to the Opposites tests and to the Performance tests. The result is that the obvious group-factor common to the Opposites tests is overlooked ; and some vaguer common factor has to be found for the three contrasted tests, in spite of the fact that they have little or nothing

[1] This may not be true of Thurstone's example, since the selection of axes is based primarily on the geometrical diagram : however, the diagram is admittedly not decisive by itself ([84], p. 169). The circular argument criticized in the text is quite common in research theses, where the writers frequently draw conclusions from the 'positive saturations,' entirely forgetting that the allocation of the positive sign has been arbitrarily made by themselves.

in common, simply because they have been arbitrarily allotted a positive sign. It would have been equally logical to have reversed the assignment: with *positive* signs for the three Opposites tests and negative signs for the three miscellaneous tests, the principle of eliminating negatives would then have prevented us from postulating any factor for the miscellaneous tests and from attempting to identify the factor common to these three with the factor common to the numerical tests. Instead we should have postulated a fourth factor specifically for the Opposites tests as such; and the overlap between factors II and III in Table 4 would have been avoided. This, I venture to submit, would be much more in keeping with the clustering revealed in Thurstone's own diagram on p. 168.

I maintain, therefore, that the somewhat arbitrary procedure by which rotations are carried out is apt to obscure lines of classification that were plain enough in the matrix of bipolar factors before rotation, and tends to import other lines of classification that are actually a product of the factorist's tacit assumptions. That in this and other instances it may be often more convenient to express the factorial composition by factors that are exclusively positive I would not for one moment deny; but, when that form of interpretation is the more appropriate, the simplest procedure, I should have thought, would be something akin to what I have called the group-factor method. In the present instance, the 15 tests, from their very nature, were obviously likely to disclose a grouping into four sharply demarcated categories. Indeed, if we look back at the initial table of correlations, we shall see that these fourfold lines of demarcation were quite clearly discernible before any attempt at factorization was made. Even if they were not, they could easily have been elicited by one of the usual devices for determining how to partition the correlation table into the necessary submatrices before the group-factor method is applied—e.g. by calculating the correlations between the rows of correlations, or by comparing the plotted contours of those rows, or finally by a rough and rapid analysis with simple summation.

Finally, let us glance at the transformation matrix (Table 3, p. 168). If this is rewritten so that the order of the rotated factors corresponds with the order of unrotated factors, the pattern I have described is readily perceived, although it is shown only in miniature (Table IV). The positive figures fall into two lines—a horizontal row, diminishing in size, and below it a diagonal ridge. To the right, there is a triangle of increasing negative figures, enclosed between the row and the ridge; to the left, a triangle of zeros (here only a single figure, unless we add a fourth factor for

opposites). This systematic arrangement makes the relation between the two forms of classification obvious at a glance.

TABLE IV

TRANSFORMATION MATRIX

Before Rotation.	After Rotation.			
	A. Verbal.	B. Numerical.	C. Performance.	Total.
I. General	·44	·30	·24	·98
II. Verbal / Non-verbal.	·89	— ·15	— ·42	·32
III. Numerical / Non-numerical	— ·09	·94	— ·88	— ·03

So far, I have confined myself to group-factors that show no appreciable overlap. In practical work, however, wherever the probable errors are low, and a close fit to the original correlations is required, it is generally necessary to allow (i) overlapping, and even (ii) occasional negative saturations in the overlapping coefficients.

(i) The use of the group-factor method need not prevent us from assuming that group-factors may overlap with each other. The overlapping factors will be dealt with along the same lines as the general factor, namely, by beginning with correlations (or residuals) uninfluenced by other factors. There are, however, limits to the amount of overlap that can be dealt with in this way with a given number of tests.

How much actual overlapping exists among the tests will, of course, depend primarily upon the way the tests have been selected. But even with sharply discontinuous groups it is hardly fair to demand that, when a factor is not an important element in a test, its saturation shall be precisely zero—that is, zero within the margin allowed by the probable error, however small that probable error may be. Is it plausible, for example, to postulate that Thurstone's verbal factor must have zero saturations for all except the conspicuously verbal tests ? Would it not be exceedingly difficult to devise a test which would exclude the smallest

trace of the verbal element, not only from the material or instructions, but even from the inner mental processes of the testees ?

(ii) Similarly, it may be doubted whether we are justified in disallowing every negative correlation, however small the probable error may be. After all, why must we so rigidly assume that ' the primary factors only act positively unless they are absent ' ? ([84], p. 71). Why should we deny that a group-factor may sometimes operate as an interference-factor ? Is it not possible that my life-long reliance on verbal methods may actually handicap me when I deal with tests of a visuo-kinæsthetic type ? Conversely, is it not possible that an interest in problems of a performance type might actually militate against an interest in numerical problems ? In short, as I tried to show in my discussion of ' negative correlations between special educational abilities ' ([35], p. 58), a psychology of intellectual abilities that is not too intellectualistic is bound to recognize that negative values among the residual correlations may be truly ' significant,' in every sense of the word.

Now, in permitting the group-factors to overlap, we shall almost inevitably be reintroducing some small degree of correlation between their saturations, and even between the factor-measurements. Our analysis thus degenerates into a ' fractionating ' type, and our selective operators are no longer ' pure.' [1] For practical purposes the oblique components thus obtained may perhaps be of more significance than the orthogonal ; the former are relatively concrete, the latter highly abstract. From a theoretical standpoint, however, it might be argued that we ought rather to modify our general factor so as to keep the secondary factors uncorrelated. And if we are to preserve the original specification of the factors so far as possible, then presumably we shall modify both, so that in the end everything will be as little changed as possible. In any case, the final outcome will be a more or less arbitrary rotation, chosen chiefly because it suits our psychological preconceptions. As an exploratory procedure, there can be no harm in such rotations : but the mere fact that they lead to an elegant and

[1] See above, p. 266.

' simple ' structure can hardly be accepted as a guarantee of their concrete psychological significance. Rather, I should argue, if the initial data have led to problems that a straightforward factorial technique is incompetent to solve, that means that the experimental stage of the research was inappropriately planned.

Finally, once we start calculating coefficients to allow for possible overlap, where are we to stop ? Unless our tests fall into abruptly separated groups, the group-factors will quickly spread out into general factors, each extending over every test : and in that case, even when dealing with residual correlations, the analysis will still have to be applied to the entire table at each step, and thus the procedure adopted will remain a ' general-factor method ' throughout. And so we come to an examination of what I have called ' general-factor methods.'

CHAPTER XIII

SIMPLE SUMMATION METHODS AND WEIGHTED SUMMATION METHODS

(*a*) *Differences in Procedure.*—General-factor methods may all be reduced to two main groups, according as they rely on 'first moments' or 'higher moments' in their attempts to fit a theoretical hierarchy to the empirical table. For purposes of practical calculation, this means that the less elaborate methods are content to use a *simple summation* of the correlations, while the more ambitious use a *weighted summation.* Spearman's well-known formula for saturation coefficients, Thurstone's centroid formula, and my own earlier summation formula, all treat the saturation coefficients as proportional to the unweighted sum or average of the corresponding columns of the correlation table. On the other hand, Hotelling's method of 'principal components,' Kelley's 'trigonometrical method,' and the 'method of least squares' in its various forms, are tantamount to requiring each row of the correlation table to be appropriately weighted before the columns are summed, *the weights in every case being proportional to the saturation coefficients themselves.*

Let us first consider what this last requirement implies (i) from a theoretical standpoint and (ii) from a practical standpoint.

(i) If we express the essential requirement *algebraically*, we are at once led to the fundamental equation which we have already encountered, namely, $R = LVL'$ or $RL = LV$; and this, it will be remembered, is the equation for the 'latent roots' (V) and the 'latent vectors' (L) of the correlation matrix R. Thus, if we put F' (the matrix of saturation coefficients) $= V^{\frac{1}{2}}L'$, and use its elements as weights, the required condition is immediately fulfilled: for, on multiplying the rows of correlations by the corresponding saturations, we have $F'R = V^{\frac{1}{2}}L' \,.\, LVL' = V.\, V^{\frac{1}{2}}L' = VF'$. If, on the other hand, we conceive the problem *geometrically*, we may

imagine our data plotted as points in a multi-dimensional frequency diagram: then the orthogonal matrix L will evidently specify the directions of the principal axes of the frequency ellipsoids—an interpretation which has suggested what Hotelling calls the 'geometrical meaning' of his procedure and Kelley's extension of the well-known trigonometrical solution of the two-variable correlation problem. Finally, I have myself shown[1] that the same equation is reached if we regard our analysis as an ordinary problem in *multiple correlation*, and solve it in the familiar way by the method of least squares: instead of seeking the best-fitting single line for a two-dimensional array of points (as we do in finding a suitable coefficient to express a simple correlation between two variables only), we now seek the best fitting r-dimensional sub-space for an n-dimensional array (r being the rank of the factorial matrix and n that of the original matrix of test-measurements). This second set of methods, therefore, consists essentially in the time-honoured *Hauptachsentransformation* that constantly crops up in the solution of so many mathematical and physical "problems of best fit," and now reappears in slightly varying guises.

(ii) As regards practical computation, the requirement by its very nature at once suggests some form of progressive approximation: thus, we might first try to guess the approximate saturations as nearly as we can; and, taking the guessed figures as trial multipliers, calculate the saturations resulting from these; then, if the results do not tally with the initial weights, we could repeat the process until they do. The 'iterative procedure' proposed by Hotelling [79], the successive 'rotation of axes' proposed by Kelley [85], and the process of 'repeated substitution' employed by Rhodes [93], all follow some step-by-step procedure such as this. Now, my own contention is that if, instead of starting with any plausible guess (as, for example, Hotelling[2] appears to do), we start, as it were, from scratch, putting the weights all equal to unity, and then mechanically calculate and recalculate the approximate

[1] [93], pp. 247 f.; cf. above, p. 164.

[2] "Round numbers may be chosen at the beginning, and the process will converge to the correct values anyhow" ([79], p. 431). I do not suggest that my mechanical procedure is quicker than Hotelling's, when only two or three decimal places are required. It is, however, easier for the beginner, who has no experience to guide his guesses; and, as explained below, the mechanical procedure followed by the construction of a geometrical progression seems speedier and more accurate if highly exact figures, running to many decimal places, are needed (Miss Dewar and I, for example, used it, when we desired to show that correlating persons and traits leads to arithmetically identical figures).

multipliers by the same regular procedure, the successive improvements will form a systematically converging vector-series whose first term will consist of the values given by simple summation and whose limit can be directly determined ([93], p. 287, [102], p. 188).

Once again the advocates of each of the two main methods are highly critical of the other's procedure. Thurstone, for example, avers that the " method of principal axes " cannot give " psychologically meaningful results," and that, " so far as the psychological problem is concerned, such a solution is not acceptable " ([84], pp. 120, 130). Kelley replies that Thurstone's procedure has " the weakness of an arithmetic average of semi-disparate things," and maintains that " the logical foundations of the principal axes method and the centre of gravity method are irreconcilably different " ([85], p. 62).

(*b*) *Relations between the Saturation Coefficients.*—But once we have observed that the figures obtained by the two methods simply represent different stages in the same sequence of approximations towards one and the same final values—the values, namely, obtained by the direct solution of the ' characteristic equation '—a reconciliation becomes quite easy ([95], pp. 288, 291). The approximations themselves can be regarded as proceeding by a repeated self-multiplication of the correlation matrix—as a result of correlating and recorrelating, as it were, the columns of correlations already obtained. And this procedure, as I pointed out in an earlier article, will ultimately reduce any empirical correlation matrix to a matrix of rank one, i.e. to a perfect Spearman ' hierarchy.' Thus, " as we keep on correlating, the sums of the products approximate more and more closely to the final proportionate values of the saturation coefficients : Thurstone merely takes the first set of product-sums ; the values obtained by Kelley's method (and Hotelling's) are those reached by carrying the same principle several steps further " ([102], pp. 188–9).

Since this interpretation has also been questioned, I propose to restate a little more fully, first in algebraic and then in arithmetical form, the argument on which it is based. In an earlier chapter it

was suggested that any matrix of correlations or covariances could be regarded as a sum of 'hierarchies,' i.e. of matrices of rank one. In matrix notation this structure can be clearly expressed by what I have called the canonical expansion of any correlation matrix: viz.:

$$\begin{aligned} R &= v_1 E_1 + v_2 E_2 + \ldots + v_n E_n \\ &= H_1 + H_2 + \ldots + H_n, \end{aligned}$$

where v denotes the factor variances numbered in descending order of magnitude, E_j 'reduced or unit hierarchies,' and H_j the series of one-rank matrices obtained by post-multiplying the vector of saturation coefficients for the jth factor by its transpose in the ordinary way. The 'unit hierarchies,' it will be remembered, possess two important properties characteristic of selective operators [115]: viz. (i) $E_j^m = E_j$; (ii) $E_i E_j = 0$, $(i \neq j)$. Now, unless n (the number of tests) is very small, the determinants required for the evaluation of the latent roots and vectors are much too large for explicit calculation. But from these two properties it follows that:

$$\begin{aligned} R^m &= v_1^m E_1 + v_2^m E_2 + \ldots + v_n^m E_n \\ &= v_1^m E_1 \text{ (approx.)} = v_1^{m-1} H_1, \end{aligned}$$

provided m is taken large enough to make the ratio v_2^m/v_1^m (and *a fortiori* the ratios $v_3^m/v_1^m, \ldots, v_n^m/v_1$) negligible. R^m can be easily computed. We have then merely to add up its columns; and thus, *by applying 'simple summation' at a higher stage*, we can at once obtain values closely proportional to the 'saturation coefficients' or 'factor loadings.' The error incurred by taking R^m in place of R^∞ depends primarily on v_2^m/v_1^m, and will therefore diminish almost in geometrical progression: the smaller the size of m, the larger the amount of error. Evidently, therefore, the figures obtained for the factor loadings by giving m the smallest possible value, namely 1 (i.e. summing the columns of R just as they stand), form the first and simplest approximation to the figures that would be obtained by taking $m \to \infty$ or by attempting a direct solution. The figures reached by Hotelling's method and by Kelley's are virtually equivalent to those obtained by summing R^m, where m is still not infinitely large; hence we must regard these values too—as their advocates admit—as being equally approximations, though doubtless much closer than those derived from the initial R^1.

It follows, therefore, that by repeated self-multiplication we can in theory reduce any matrix of correlations or covariances to a matrix of rank one.[1] In practice there

[1] The self-multiplication of a determinant or matrix (really a special case of the familiar root-squaring device) has long been in use by physicists and

would seem to be two alternative ways of approaching this result (both briefly indicated in my previous article ([101], p. 91). Their equivalence may be proved as follows.

Let w_o denote the summation operator [1, 1, ..., 1]. Then the simple unweighted sums of the columns of R may be denoted by w_oR, a row-vector like w_o itself. Similarly, the sums of the columns of R^4 (say) may be written (i) $w_o(R \times R \times R \times R)$; but this is clearly the same as (ii) $(w_oR)\ R \times R \times R$.

The identity suggests two different working procedures; (i) we may *begin by multiplying* the matrix R by itself: repeat the process m times; and then end by adding the columns of the final product-matrix; or (ii) we may *begin by adding* the columns of R, multiply the whole matrix by this row-vector, i.e. weight each row of correlations and add once again; then, if we repeat the multiplications m times, the final sums will obviously be the same as before. We may term (i) 'table-by-table multiplication' and (ii) 'table-by-column multiplication.'

Hotelling's iterative method might be described as a telescoped form of (ii), and his matrix-squaring method as a curtailed form of (i). By taking guessed approximations at each step instead of exact multipliers, the final result is more rapidly reached than if the full figures are retained at every step; but at the same time the underlying nature of the process is obscured. When the computer carries out the calculations in full, he quickly perceives that the differences between the successive approximations are each the same fraction of its predecessor, or nearly so: hence there is no need to

mathematicians for solving characteristic equations and the like. My late colleague, R. C. Howland, for example, has described it in his *Note on a Type of Determinantal Equation* [62] in connexion with the solution of certain dynamical problems in physics. I do not think, however, that it had been explicitly shown before that the ultimate result would invariably be a 'hierarchy': a proof without matrix notation was given in [102], p. 183.

Since, by the product-moment formula, $R = MM'$ (where M is the initial matrix of measurements), the columns of R (which in its most general form may consist of covariances, averaged or unaveraged, adjusted or unadjusted) may be described as comprising the 'first moments,' and the columns of $R^m = MM'MM' \ldots MM'$ as containing 'higher moments' (cf. [93], p. 287; [102], p. 178). Hence these terms have occasionally been used to distinguish the two methods.

continue the iterative process. He can easily extrapolate by summing the geometrical progressions.[1]

The relations between the various results will become obvious even to the non-mathematical if we glance at a concrete example. I choose the table of correlations factorized in two different ways by Kelley ([85], p. 58, Table IX) to demonstrate the 'irreconcilability' which he believes to exist between Thurstone's results and his own.

In Table V the figures at the foot of the three columns show the results obtained by a first summation of the correlation matrix R taken just as it stands, i.e. they form the vector w_0R. Now, adopting the table-by-column method, we take these sums (w_1, say) as weights for multiplying the original correlations, row by row, and then make a second summation to obtain the vector $w_1R = w_0R^2$ (cf. second line of Table VI). To render the comparisons visible to the eye, the weights are reduced at each stage to fractions of unity: the first reduction is shown in the last line of Table V.

TABLE V

1st Multipliers.	Correlations			Totals.
1	1·00	·70	·26	1·96
1	·70	·75	·45	1·90
1	·26	·45	·35	1·06
4·92)1·96	1·90	1·06		4·92
	·398374	·386179	·215447	1·000000

[1] The summation of the geometrical progression to shorten the labour of successive approximation was briefly described at the March meeting of the British Psychological Society (1937): full working instructions are given in *Notes on Factor-Analysis* (1936) obtainable from the laboratory, from which both the foregoing proof and the following example are quoted (cf. also Appendix I, below). In the *Notes* it was not made sufficiently clear that, in general, these elaborate methods of computation are only necessary when figures are required to several significant places (e.g. for comparing results by different methods, as here or in [114], or again for analysing a table of covariances as distinct from correlations). When the variances are unknown, and estimated communalities have to be inserted in their place, it would be absurd to employ so refined and laborious a procedure. In my own experience, the resulting improvement is, as a rule, comparatively small ([93], p. 294, [128], p. 63). Davies, however, has recently found that, in several tables of correlations between persons, the simple summation method may leave residuals which the investigator judges to be significant, whereas, if the least-squares method is used, the residuals are at times considerably reduced, so that nothing but a general factor can be validly demonstrated (cf. [130]).

The weighting and summing is repeated again and again, until, after 9 or 10 repetitions, the final results yield no change, at any rate so far as the sixth digit in each number (Table VI); it will be noted that the differences between corresponding numbers in the successive lines become smaller and smaller, and that approximately

$$\frac{.415804 - .398374}{.419210 - .415804} = \frac{.419210 - .415804}{.419869 - .419210} = \ldots = \frac{1.75782}{0.33950} = \frac{v_1}{v_2},$$

and similarly for the second and third column. Moreover, the figures thus obtained by table-by-column multiplication will be found identical with those obtained by table-by-table multiplication ([8], p. 185).

In the present example, since the determinants of the correlation matrix are small, a check is procurable by undertaking a direct solution. The figures so obtained are inserted in the last line but one, and coincide with the values that would be reached by increasing the index of R indefinitely, as estimated by summing the three geometrical progressions.

TABLE VI

Stage of Approximation.	Saturation Coefficients. (Proportionate Values for Successive Approximations).			Factor Variance.
Summing R (Thurstone)	·398374	·386179	·215447	1·742923
„ R^2	·415804	·381779	·202397	1·754932
„ R^3	·419210	·380921	·199870	1·757263
„ R^4	·419869	·380751	·199380	1·757713
„ R^5	·419997	·380717	·199285	1·757800
„ R^6	·420022	·380711	·199267	1·757816
„ R^7	·420025	·380710	·199264	1·757820
„ R^8	·420026	·380710	·199263	1·757820 +
„ R^9	·420027	·380710	·199263	1·757821 −
. . . . .	. . .	. . .	. . .	
„ R^∞	·42002715	·38070993	·19926295	1·75782072
By successive rotations (Kelley)	·420273	·380669	·199058	1·75783

The first line of the table is identical with the figures calculated according to Thurstone's centroid method; the last line but one gives figures we should reach by Kelley's method or Hotelling's, if their rotations or iterations were continued indefinitely. The results obtained by Kelley himself with his 'new method of analysis' are appended in the last line

of all. It will be seen that the latter, too, form only an imperfect approximation; they are certainly much closer than those which are got by Thurstone's method (as calculated by Kelley), but tend towards over-compensation.

Whether our data permit us to discover one significant factor only ($r = 1$, as in Spearman's deduction of g) or whether we decide to seek for a larger number ($r > 1$), the method of analysis which I have here described *will always yield the hypothetical correlation matrix of rank* f *which best fits the actual correlation matrix (total or residual) empirically given.* Thus, the essential feature of the 'simple summation' method as compared with that of the 'least squares' is that the former is content with a matrix of rank f that yields a slightly poorer fit—in fact, as we have seen, with the first set of values, instead of the last, emerging in one and the same sequence of successive approximations. Moreover, with the least-squares method the factors are fully independent in that, *not the factor-measurements only, but also the factor-loadings are uncorrelated* (Thurstone's factor-loadings nearly always show a low correlation). Hence, where further transformations are required, the algebra and arithmetic become much simpler. Finally, if we construct an artificial correlation table from a given set of saturation coefficients obeying these conditions, and then factorize it, we shall with this method get back to the original saturation coefficients: this result is not obtainable with Thurstone's procedure or Spearman's, nor (to judge by his own example) with Kelley's.

Two defects are often attributed to the method of least squares. First, it is said that, unlike the factors reached by simple summation, those deduced by the method of least squares cannot be rotated: for, if they are rotated, they lose their essential quality of being 'principal components,' i.e. of being principal axes of the frequency ellipsoids. But the latter property as an end in itself is useful only in special cases (e.g. empirical prediction). And, as must now be evident, since one set of factors is merely an approximation to the other, it must be just as legitimate to rotate the one as the other. What is more, with the least-squares factors, owing to the non-correlation of the factor-saturations, such rotations can be much more easily performed. Secondly, the followers of the two-factor theory have objected that with the least-squares method the factors must change whenever a new test is added to the battery. Actually this is also true with a two-factor analysis; but when the original battery involves only one significant common factor, it is easier to select the additional test so that the change is negligible. The

difficulty, however, is by no means peculiar to methods of factor-analysis as such, but to all attempts to deduce an average from a small set of data. If I calculate an average and a mean deviation from only a dozen persons, and then add a 13th and recalculate, the recalculated figures are bound to differ, unless that 13th person is very carefully picked. Similarly, if I calculate a set of factors from only a dozen tests, those factors, being nothing but an average and a set of mean deviations derived from those tests, are almost bound to alter when a 13th test is added.

In making these comparisons I have so far assumed that the same complete matrix is factorized in either case. What I have said above needs a little modification when applied to the best-known representatives of the two methods, since the different writers complete an incomplete correlation matrix in different ways. Hotelling's method inserts self-correlations of 1·00 in the leading diagonal: hence, with n tests, the correlation matrix cannot be perfectly fitted without using n factors—even if the intercorrelations by themselves are perfectly hierarchical. On the other hand, Thurstone's method (in theory, though not it would seem in practice) is applied to a 'reduced correlation matrix' in which 'communalities' are inserted; and these are so chosen as to reduce the rank of the matrix artificially to the lowest obtainable number. In such a case an exact fit can be obtained for the inter-correlations with less than n factors; for Thurstone does not count the specific factors, as they are not required to reconstruct the reduced correlation matrix. His gain, however, seems purely academic. It is true that with (say) a dozen tests the correlations could always be perfectly fitted with 7 factors only. But in how many tables are the probable errors so minute that we should look for as many as 7 factors? Moreover, by merely printing the correlations to a different number of decimal places, we may in many cases alter its rank far more drastically than by any change in the arbitrary communalities. However, as I have so often insisted, what our factors have to interpret is not so much the correlation matrix, R, as the original set of measurements, M; the device of correlating is only an incidental step, and not always the best. Now, Hotelling's method, like the method of least squares, always yields a better fit to the original measurements, even if we retain only a small number of factors (e.g. significant factors only); with n factors the fit is perfect. With the Thurstone method, on the other hand, as with the summation method, we have to invoke the specific factors as well before we can obtain so good a fit to the original measurements ([101], p. 80). And since both Thurstone and Spearman believe in

the concrete reality of specific factors, and even maximize their influence, the number of factors that their theories envisage is actually $r + n$. Such a result surely contravenes Thurstone's own postulate of parsimony. Hotelling's method tacitly assumes that, by 'correcting' the correlations or by selecting and refining the tests, the specific factors as such can be virtually abolished: alternatively, if every test is applied twice, to obtain its 'reliability' coefficient, we may regard its specific factor as convertible into a group-factor common to the two applications, which, like every other group-factor, becomes bipolar when the general-factor method is employed ([93], p. 307). In either case, economy in the number of factors is then achieved, not by an arbitrary reduction of the apparent rank, but by rejecting those factors that are statistically insignificant and maximizing the influence of those that are retained.

(*c*) *Relations between the Factor-measurements.*—These, then, being the differences between the factor-loadings or saturation coefficients obtained by the two methods, what is the difference between the factors themselves? First of all, since with the method of least squares n factors are in theory obtained for n tests, the factor-measurements for each person can be calculated exactly: this is still true, even if we do not carry the analysis to its full completion.[1]

[1] Cf. [101], pp. 80, 92–3. In the illustrative example there given, it will be observed, in order to obtain the factor-measurements for all the persons in the first factor alone (para. 8) only the 'regression coefficients' for that one factor are required (cf. para. 7), i.e. only the one latent root (the variance for the first factor) and the one latent vector (direction cosines for the first factor) are required. Thus with this method of calculation the 'regression coefficient' is simply a weighting coefficient, and the process of 'estimation' simply the direct calculation of a weighted average. Indeed, so long as we confine ourselves to statements about the particular group studied, the term estimation is a misnomer, since no reference is so far made to the possibility of error. But if we *know* on *a priori* grounds that one factor only is really operative, and that the remaining factors represent errors of measurement, etc., then we may certainly regard the factor-measurements deduced for that one factor as estimates of the true factor-measurements: in such a case, what we are doing is simply to find the hypothetical $n \times N$ matrix of rank $r = 1$ which approximates most closely to the observed $n \times N$ measurement matrix of rank n, with the principle of least squares as the criterion of best fit. If, however, we calculate factor-measurements not for one but for all the n factors, then the term 'estimation' again becomes a misnomer, so far as the group studied is concerned, because then the fit is perfect. Of course, when we

On the other hand, with the centroid method there are more factors than tests; and hence the factor-measurements can only be estimated approximately: the calculations for Thurstone's factors necessarily involve the same kind of indeterminacy as has been shown to be involved in the estimation of Spearman's *g*—and for precisely the same reason.

There is, however, a still more important difference. In my Memorandum I briefly stated that the simple summation method virtually treats the first or dominant factor ('*the* general factor') as the simple sum or average of the several test scores (just as it virtually treats the factor loadings as the sum or average of the correlations),[1] while the least-squares method treats it as given by a weighted sum or average of the several test-scores, determining the weights on the lines of the ordinary multiple regression equation. In the same way, the secondary or subdominant factors are treated as unweighted and weighted averages of the deviations.

These corollaries may perhaps best be demonstrated as follows. As before, let M be the matrix of measurements, so standardized that $MM' = R$, the matrix of correlations or standardized covariances. Let F be the 'factorial matrix' containing the saturation coefficients, and P the 'population matrix' containing the factor-measurements. Then $M = FP$ and $P = F'R^{-1}M$ ([101], p. 80). For simplicity consider only the first or 'general' factor (g), i.e. the first column of F (f_1) and the first row of P (p_1). Then, by the centroid formula:

go on to treat the group studied as a sample only, the factor-measurements deduced exactly for this group may be regarded as approximate estimates for the entire population. If the group has been tested twice, so that the 'reliability' of each test is known, that can easily be taken into account in weighting the tests: we have only to multiply each test (weighted as above) by the square root of its reliability coefficient.

[1] [93], p. 287 and pp. 247–8, 300; cf. [102], p. 176. As noted in the latter paper, Kelley in [85] has since independently reached much the same conclusion: in particular he shows that, if a correlation matrix for two variables only be analysed by Thurstone's method, the first factor proves to be merely an average of the initial variables: he has "not undertaken an algebraic analysis of Thurstone components for 3, 4, or a larger number of variables"; but in three specific cases his arithmetical solution manifestly leads to a similar result ([85], pp. 58–61).

$$f_{jk} = \frac{\Sigma r_{ik}}{\sqrt{\Sigma\Sigma r_i r_i}} \text{ i.e. } f_1 = \frac{w_o R}{\sqrt{w_o R w_o'}}$$

where w_o as usual denotes the simple summation operator. Accordingly, $p_1 = f'_1 R^{-1} M = \dfrac{w_o R R^{-1} M}{\sqrt{w_o R w_o'}} = \dfrac{w_o M}{\sqrt{w_o M . M' w_o'}}$

where the numerator is merely the unweighted sum of each person's measurements and the denominator reduces these sums to unitary standard measure.

The foregoing argument implies that the same correlation matrix, $R = MM'$, is used throughout, whether we are calculating p or f. Now for the former purpose—'the appraisal of abilities' (p)—Thurstone takes an R having diagonal elements of unity ([84], p. 227); for the latter purpose—the calculation of factor-saturations (f)—he proposes to substitute a different R, containing in its diagonal the highest correlation in each column (pp. 89, 108). In that case the expression RR^{-1}, in the final equation for p, cannot be taken as *precisely* equivalent to the unit matrix, I; the change, however, only involves a slight difference in weighting.[1] Thurstone himself

[1] In the unabridged version of my *Memorandum* I gave the proof in a more general and a somewhat more elaborate form, which shows that the result is not limited to unweighted summation nor to a correlation table with the self-correlations taken as unity. With the various criticisms of my conclusion I shall briefly deal in Part III. Here I limit the proof to the simplest possible case, not only because the underlying principle will then be clearer and, I hope, beyond all controversy, but also because this simpler proof alone is needed to justify my use of this procedure in calculating such factor-correlations in the past and in deducing certain results in the chapters that follow. The tables in Part III (pp. 391, 398–9) will serve to illustrate the present argument by a concrete instance; Kelley's first example (*loc. cit.*, p. 58, Table IX) will serve to illustrate the extension of the same conclusion to cases where communalities are inserted in the leading diagonal instead of unity.

In general, however, if the correlation table is small, and if the hierarchical tendency is steeply graded (and therefore not overlaid by marked group-factors), the saturation coefficients obtained respectively "with unity" and "with communalities in the diagonal cells" differ in the way illustrated in Table IIIA and IIIB: cf. [93], p. 307. Thus we may convert the $n \times 1$ factorial matrix given in Thurstone's Table 5 ([84], p. 101) into a 'prop-ladder pattern' by appending the coefficients for the specific factors as deduced from Spearman's two-factor theorem; the resulting $n \times (n+1)$ factorial matrix will reproduce his Table 1 quite as well as the $n \times n$ matrix shown in his Table 2. The factorial matrix reached by Godfrey Thomson (*J. Ed. Psych.*, XXV, p. 367) on applying Hotelling's original method to a

actually states : " fortunately the diagonal entry may be given any value between zero and unity without affecting the results markedly " (p. 108).[1]

We may, therefore, fairly conclude that, in principle, the *saturation coefficients given by the centroid equation are virtually the correlations of those tests with the unweighted average of their scores*, and that, if there is any slight differential weighting, it will depend solely on the somewhat arbitrary figures chosen for the communalities, and in any case will not " affect the results markedly," except when the tables are small.

The factor-measurements obtained by the method of least squares, on the other hand, are weighted averages. But there will be no need to apply the principle of least squares afresh or to calculate the ratios of the resulting determinants : for we now have $F'R^{-1} = V^{-1}F'$. Thus we have now merely to divide each column of saturation coefficients by the corresponding factor-variance, and we at once obtain the appropriate weights.

pure hierarchy with unity in the diagonal cells is also of the same quasi-triangular type, with nearly all the peculiar features illustrated above in Table IIIB. In both examples it could be shown that the factorial matrix obtained with the one method could be rotated to give that obtained with the other by a quasi-triangular transformation matrix of the kind described on p. 305. As will be seen by considering its special features, *when the correlation table is very large, this transformation matrix tends to become a unit matrix with a column of units prefixed.*

[1] The student who does not follow the above proof by matrix algebra can satisfy himself of the equivalence either by studying the tables below (pp. 391 f.) or by following the inverse procedure. Let him first calculate the factor-measurements for *g* by summing the standardized measurements for all the tests, and correlate these sums with the measurements for any one test : the correlation of each test with *g* will then prove to be identical with its saturation coefficient for *g* (see [101], p. 75). If he likes, he can now generalize this algebraically by using the formula for the correlation of sums ([47], p. 197, eq. 147, putting $a = 1$).

It may be noted that we have here a method of reconciling Thomson's sampling theory with Spearman's two-factor theory : for we have only to take the sum of the test-measurements to be expected in accordance with the sampling theory, and correlate these sums with the most probable measurement for any one test : we then find that these correlations have the same properties as Spearman's saturation coefficients as deduced from a pure hierarchy. So that once again the sum of the test-measurements has the properties of *g* and *vice versa*.

CHAPTER XIV

TESTS OF SIGNIFICANCE AND OF HIERARCHICAL TENDENCY

BEFORE accepting a set of mathematical factors we require to establish the statistical significance of the variances and saturation coefficients that specify them. This involves a number of somewhat neglected and difficult questions. To attempt a formal and technical inquiry into the whole problem would be out of place in a discussion intended for the general student. I shall content myself with comparing some of the more obvious and more familiar proposals, and shall endeavour to indicate in the broadest way how they are related to each other and to the methods adopted in general statistics.[1] For the most part I shall assume that the factors under discussion are the factors immediately resulting from an analysis by weighted summation—i.e. are the factors specified by the latent roots and vectors of the correlation matrix: for with this method the factors and their saturations are independent of each other, and so lend themselves to the simplest form of demonstration. But, as we have already seen, many writers prefer to convert these factors forthwith into another set having different values and possessing, as they believe, a psychological meaning which the first set cannot claim. First of all, therefore, it may be well to consider what general changes such a conversion is likely to entail.

[1] My review does not profess to be complete. Kelley's interesting discussion of the problem ([85], pp. 10–17) relates primarily to the steps in his own rotation-method of factorization, but could be brought into line with what follows. Similarly, Hotelling ([79], p. 437) gives a method for testing the 'reality' of his components (which are obtained from corrected correlations) based on the reliability coefficients: but here I have confined myself throughout to correlations obtained from a single application of the tests. Thurstone also suggests 4 or 5 supplementary criteria ([122], pp. 65–6) in addition to the two I have discussed below.

Effect of Rotation on Factor-variances.—According to Thurstone, when factors have been extracted by the summation method, they must always be transformed, by a rotation of the reference axes, to factors which (i) have exclusively positive saturations and (ii) have as many zero saturations as possible. Now, if the method of weighted summation has been used (and the same is true of the method of simple summation so far as it approximates to that of weighted summation), the first factor extracted accounts for the largest amount of variance that could possibly be accounted for by a single factor alone; the second factor accounts for the largest possible amount of the residual variance; and so on. Again, on taking the factors in the opposite order, we find that the last factor accounts for the smallest possible amount of the variance, consistent with the fact that all the variance has to be accounted for by n factors at most; and so on for the other factors. Thus, the set of factor-variances obtained by this procedure has a wider range and a larger standard deviation than any other set that can be derived from the same table. Consequently, the substitution of any other set of factors by rotation is bound to reduce the inequalities of the factor-variances, and so to diminish their standard deviation. Moreover if, in accordance with the requirements of 'simple structure,' the rotation is so arranged that each of the new factors has a large number of zero saturations, and presumably about the same number ("at least r zeros," where r is the number of factors [84], p. 156), then the ultimate result must obviously be to *flatten out the differences*. Indeed, a single glance at Thurstone's tables, derived from the same data before and after rotation of the factors (e.g. [84], pp. 108, 169, [122], pp. 113–6), is enough to show that this levelling out is the most conspicuous change. Thus, the attempt to obtain a 'simple structure' obscures one of the most characteristic points of the given correlation table, namely, the particular degree to which it is dominated by one or two outstanding factors.

Almost all the correlation tables met with in psychology are distinguished by the fact that the standard deviation of their factor-variances is exceptionally high. This means

that they display simple and regular patterns. An extreme example is the covariance matrix whose pattern is hierarchical. Since in such a case the variance of the one 'general' factor is equal to the total variance and that of all the others equal to zero, the standard deviation is a maximum. According to Spearman, a hierarchical pattern of this type is the special feature of every correlation table obtained with cognitive tests.

Nevertheless, that is not what we should have anticipated. Many psychologists—Thorndike, for example—apparently expected the mind to be composed of a number of independent 'fundamental abilities' or 'faculties,' whose influence was approximately equal.[1] Indeed, this, it would seem, is the assumption which underlies the hypothesis of 'simple structure.' But on this assumption (provided the tests for each ability were equal in number and equally efficient) we should find, even with the method of weighted summation, that the factor-variances were all of much the same size. In that case both their range and their standard deviation would be approximately nil.

The 'Principal Tetrad-difference' Criterion.—In practice what we usually discover is a set of factor-variances whose standard deviation lies between these two extremes, sometimes inclining more towards the maximum value, more rarely tending in the direction of zero. In the early days of factor-psychology it was assumed by many writers that any departure from the hierarchical ideal was attributable to errors of sampling; and Spearman's examination of all the available correlation tables [24] provoked a sharp controversy as to whether a single general factor was or was not sufficient to explain the entire amount of observed correlation, and in particular whether his criterion—the calculation of the intercolumnar correlations—provided a safe and adequate test. Later, he himself suggested a second and still more elaborate criterion, namely, the calculation of all the 'tetrad differences' and their sampling errors, with a view to showing that, within the margin of error, the 'tetrad differences' are all zero [52]. Our first task, therefore, will be to examine these two criteria, and to see

[1] *Educational Psychology*, 1903, p. 39.

how they are related to each other and to the more important alternatives that have been proposed. Let us begin with the tetrad-difference criterion.

If we regard the correlation table as a matrix or determinant arising from the solution of a set of linear equations,[1] this criterion can be reduced to a very simple form. As has often been pointed out, it is—under a different title—the regular method for demonstrating that a determinant is of unit rank.[2] The great objection to its use is the time and labour involved: with a dozen tests the intercorrelations would yield 1,485 tetrad differences to calculate (not, however, all independent); with Thurstone's latest table [122] over a million. But when the determinant is symmetrical, it is in theory unnecessary to compute *all* the tetrad differences: it is sufficient[3] to show that the *principal* tetrad differences—i.e. the diagonal minors of order 2—are all zero. The sum of these diagonal minors, $\Sigma\Sigma\begin{vmatrix} r_{ii} & r_{ij} \\ r_{ij} & r_{jj} \end{vmatrix} = (\Sigma r_{ii})^2 - \Sigma\Sigma r_{ij}^2$. If, therefore, the correlation matrix is of unit rank, the sum of the squares of all the correlations (inter- and self-) must be equal to the square of the sum of the self-correlations, i.e. squares of the saturations with the single factor: in other words, the standard deviation[4] of the correlations should be equal to $\frac{1}{n} \times$ the total hypothetical test-variance. This hypothetical total can be determined with sufficient accuracy by the methods described in the appendix; and the probable error of the standard deviation can be estimated in the usual way. It is curious to note that this form of the tetrad-difference criterion depends on calculating precisely those tetrad differences which the ordinary form of the criterion omits.

Since $\Sigma r_{ii} = \Sigma v_i = \Sigma v_j$ (where v_i and v_j denote the n test-

[1] Cf. *Marks of Examiners*, p. 247.

[2] The 'tetrad' is a two-rowed minor determinant renamed; and the 'tetrad difference' its expansion. Hence the criterion may be regarded as a special application of the familiar property of determinants, namely, that 'if the corresponding elements of two columns (or rows) are proportional, the determinant vanishes.' For the criterion as a theorem in matrix algebra, see Cullis [26], 1918, II, pp. 93, 139; and cf. Bocher [15], 1907, p. 34 f.

[3] Strictly we ought also to show that the diagonal minors of order 3 are also zero. But this is needed only to rule out certain patterns involving negative signs which are never likely to arise in psychological work.

[4] The 'unadjusted' standard deviation, if we are dealing, not with a bipolar matrix of residuals, but with an observed table of positive correlations.

variances and the n factor-variances respectively) and $\Sigma\Sigma r_{ij}^2 = \Sigma v_f^2$, the hierarchical requirement is equivalent to demonstrating $\sqrt{\Sigma v_f^2} = \Sigma v_f$, which is obviously satisfied if all the factor-variances except the first, i.e. $v_2, v_3, \ldots, v_n, = 0$. The expression itself suggests the possibility of *a simple criterion in terms of the standard deviation of the factor-variances.*

The True Hierarchical Issue.—But the whole issue, as it seems to me, should now be formulated rather differently. The parties to these controversies often write as though there were only two extreme alternatives: either to explain all the variation by the smallest conceivable number of factors (namely, one, in Spearman's case) or to explain it by the greatest conceivable number (which they take to be n, the number of correlated tests). Rather, it would seem, the alternatives lie between (i) seeking a factorial matrix that will exhibit the variance as mainly concentrated in a few dominant factors of varying importance and (ii) accepting a factorial matrix that will exhibit the variance as distributed almost equally among a large and indefinite number. And so far as concerns the first or dominant factor—'*the* general factor,' as it is commonly called—our task is not to demonstrate that it will account for everything, but simply to discover *how much* it will account for.

Save for exceptional cases in which the tests have been specially selected, there must *always* be more non-specific factors than one. Consequently, unless the sample tested is so small that the issue cannot really be decided, the answer to the tetrad-difference criterion must be the same in every case: namely, no empirical correlation table forms a perfect hierarchy. Thus, the crucial issue has been wrongly conceived. The question to ask is not, is this empirical table a hierarchy or is it not? but, how strong is its *tendency* towards the hierarchical pattern? And in theory the same inquiry should be made, not only about the initial matrix, but about each of the residual matrices (so long as they are significant) in turn. The most obvious mode of approach therefore will be this: having extracted a complete series of factors in order of their maximal contribution to the total variance, to inquire: first, how many of these factors

can claim statistical significance ? and, secondly, what is the relative importance of each one ?

The Significance of the Factors.—The first question is most simply answered by testing the significance of the individual correlations or residuals on which each factor is based. If a residual is three times the probable error, it is customary to assume that it is significant of an additional factor. There are, however, a number of familiar pitfalls attending this criterion.

First, the probable error should be that of the observed correlation : in many investigations the residual is treated as itself an observed correlation, and its probable error taken direct from the usual table ; in others it is calculated as a probable error for the difference between two observed correlations.[1] Secondly, the number of residuals should be taken into account : if there is only one residual of the size required in a table for 7 tests, or if there are half a dozen of that size in a table for 16 tests, we must not thereupon conclude that they are significant merely because each reaches the conventional level : for with chance distributions these are just about the numbers we should expect. On the other hand, if there is a large number of residuals somewhat less than three times the probable error, we must not thereupon conclude that there is no further factor : we cannot even conclude that a further factor is improbable : for, if we are dealing with a single coefficient and that coefficient is, say, only twice the probable error, the chances are still more than 4 to 1 against so large a figure arising as a result of random sampling ; and if half the figures in the table are of this order, their cumulative evidence will be strongly in favour of a common factor, even though none of them is up to the conventional level.[2]

[1] The residual is the difference between an observed and a given hypothetical correlation, and the hypothetical correlation is not to be treated as if it were another observed correlation subject to sampling errors. The proper procedure is that prescribed for testing " the significance of the deviation of an observed coefficient from the expected value " (e.g. Fisher, [50], pp. 189–90, ex. 31).

[2] In the early investigations with mental and scholastic tests, some of us, who found certain striking deviations from the expected hierarchical value occurring in table after table, claimed them as at least highly suggestive of group-factors. Our critics, on the other hand, observing that the residual or 'specific' correlations were seldom equal to three times the probable error, cited the same figures as disproving the existence of such group-factors, even when the odds were in fact definitely against such figures having arisen as

It is best, therefore, wherever the issue is crucial, to indicate the probabilities for and against the chance hypothesis. The most obvious procedure is to take the ratio of each residual correlation to the standard error of the observed correlation from which it has been computed, and then to compare the actual distribution of these ratios (or its standard deviation) with the theoretical distribution that would be expected if their fluctuations were due to chance (the standard deviation of this distribution will, of course, be unity). The probability that divergences as great as this, or greater, would arise from random sampling can then be estimated in the usual way. Better still, we may calculate $\chi^2 = \Sigma(z - z_0)^2$ $(N-3)$, where z_0 = the expected value of z—e.g. $\tanh^{-1}(r_{ig}r_{jg})$ if only a single general factor, *g*, has been extracted—and enter Fisher's P-table ([50], p. 110) with $\{(\frac{1}{2}n(n-1) - sn + \frac{1}{2}s(s-1)\}$ degrees of freedom, where n = number of tests and s = numbers of factors so far extracted (cf. p. 463).[1]

errors of random sampling. For example, in my 1917 *Report* ([35], p. 59) and elsewhere I found evidence for a special 'verbal' or 'linguistic' factor. In *The Abilities of Man* (p. 237), however, Prof. Spearman examines the evidence for or against 'the assumed special ability for verbal-abstract operations' and cites the 'decisive' work of Davey [54]. If I understand his figures rightly, the tetrad difference is apparently only 1·4 times its probable error. Here, therefore, the evidence for a verbal factor certainly fails to reach the level required for statistical significance: at the same time the odds are still *against* rather than in favour of the appearance of an additional factor being the mere effect of chance. (And the later work of Kelley and Stephenson now seems unquestionably to confirm the hypothesis of a verbal factor.)

Because "there is no *significant* tetrad difference" we cannot infer that "there is therefore no specific correlation or group-factor," particularly when the odds are actually against the tetrad difference having arisen from chance. We can only infer that the specific correlation, though probably genuine, is not fully conclusive. To prove that a piece of evidence is not significantly positive is not to prove that it is significantly negative. Nevertheless, this fallacious argument still constantly crops up. Hence it is worth reminding the student that there are, not two opposed alternatives, but three, with the lines but vaguely drawn between them: (i) the size of this figure is consistent with its having probably arisen by chance; (ii) the size of this figure is inconsistent with its having probably arisen by chance, but is consistent with its being due to a special factor; (iii) the size of this figure is consistent with either hypothesis, i.e. the sample is too small to enable us to reach a decision.

[1] The *z*-transformation is scarcely needed with coefficients below ·35. These methods were employed in a succession of investigations by research students working at the London Day Training College, and seemed to give satisfactory results. Strictly speaking, the residuals obtained from one and the same correlation table cannot be regarded as entirely independent,

Instead of comparing *each* residual with the corresponding error (as in the two preceding methods), the authors of many research theses are content to compare the absolute *mean* of the residuals (or their root-mean-square) with the mean deviation (or the standard error) to be expected by chance. Usually, the latter is computed from the probable or standard error either of the mean correlation [1] or (less usually) of a zero correlation,[2] with a sample of the size observed. But I agree with Guilford that, for exact purposes, " this criterion is too crude." [3] Not infrequently it leads to the extraction of too few factors. The standard deviation of the residuals may be only equal to, or even less than, the standard error of the mean correlation; yet several isolated residuals may be more than twice the standard error of the corresponding correlations.

The Measurement of the Hierarchical Tendency.—The relative importance of the several factors (*a*) in each of the tests is specified by the squares of saturation coefficients and (*b*) in the matrix as a whole is specified by the factor-variances, i.e. by the sums of those squares. In an earlier paper I have shown that, by repeated self-multiplication or ' squaring,' any matrix, if not already hierarchical, will be reduced sooner or later to a hierarchical form as nearly perfect as we desire ([102], p. 186). When this stage is ultimately reached, the figures in the leading diagonal will be proportional to the squares of the saturation coefficients for the first factor. The speed with which a hierarchical pattern is thus approached depends essentially on the amount of separation between the first latent root and the

even when due to no common factor: for the errors of correlations are themselves correlated; and, if some of the errors are large, the consistency conditions that must necessarily obtain may of themselves produce some slight hierarchical tendency. To take this into account would require an elaborate correction of the ordinary formula; but in most cases the correction is of an order that does not seriously affect a broad determination along the simpler lines. See M. Davies, ' The General Factor among Persons ' (unpublished appendix to thesis): cf. also Pearson and Filon, ' On the Probable Errors of Frequency Constants and the Influence of Random Selection on Correlation,' *Phil. Trans.*, CXCI, A, pp. 229–311; and Pearson ' On Lines and Planes of Closest Fit,' *Phil. Mag.*, II, p. 559 *et seq.* (which deals more particularly with the fitting of principal axes in the case of more than two variables).

[1] Cf. [84], p. 147.

[2] Cf. [107], p. 45.

[3] *Loc. cit.*, p. 495.

rest. Obviously if the first latent root—i.e. the factor-variance for the first factor—is equal to the total test-variance and all the other latent roots are zero, the amount of separation is a maximum. We may, therefore, assess the hierarchical tendency of any given matrix by comparing the figures in its leading diagonal with those in the leading diagonal of its square or higher power.[1]

Let f_{i1} (= r_{ig} in Spearman's notation) denote the saturation of the ith test with the first (or 'general') factor (g), and similarly for the other factors. Let $v_1 = \Sigma f^2_{i1}$ denote the factor-variance for the first factor (i.e. the largest 'latent root' of the correlation matrix), and similarly for other factors. And let trR denote the 'trace' of the matrix R, that is, the sum of the elements in its leading diagonal (i.e. of the 'self-correlations,' 'test-variances,' or 'communalities,' as they are variously called, when R is a correlation or covariance matrix.) Then, if R is hierarchical, $trR = \Sigma f^2_{i1}$; and, on squaring and re-squaring this hierarchical matrix, we shall have

$$\frac{\Sigma f^2_{i1}}{trR} = \eta_1, \frac{tr(R^2)}{(trR)^2} = \eta_2, \frac{tr(R^4)}{(trR)^4} = \eta_4 \text{ (say)}, \ldots$$

all equal to unity. If, however, R is not hierarchical, any one of these 'trace-ratios' may still be used to measure its hierarchical *tendency*, which we can now define as the speed with which self-multiplication produces the hierarchical form. Each of the ratios, as I hope to show, can be given an easily intelligible interpretation; and each has its own special advantages in answering a special form of the fundamental question.

Adopting the terminology proposed in a previous paper the three ratios may be called the criterion of first, second, and fourth moments respectively.[2] Now the use of second moments, as is there explained, is equivalent in principle to the well-known device of correlating the columns of correlations. And since this is a more familiar conception I shall endeavour to elucidate the special merits of each of these three criteria from that particular standpoint,

[1] A more rigorous but technical proof is given in [115], p. 162.

[2] [93], p. 287, footnote 1; [102], p. 178.

relying rather on suggestive analogies than on technical proofs to make the various possibilities intelligible to the non-mathematical student.

The Intercolumnar Correlation as a Hierarchical Criterion.—According to the definition of a hierarchy accepted here,[1] a table is hierarchical if its rows (or columns) are all proportional to one another. With this definition, the completeness with which a given set of coefficients tends to the hierarchical form can be judged by the '*unadjusted* correlations' between its rows (or columns). If the coefficients in the matrix are distributed entirely at random,[2] then we might expect the correlations between every pair of rows (or columns) to be approximately zero; if, on the other hand, the matrix is perfectly hierarchical, all the rows will be in perfect correlation with each other, either positive or negative: that is, in matrix notation,

$$D^{-\frac{1}{2}}RRD^{-\frac{1}{2}} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{bmatrix} \text{ or } \begin{bmatrix} 1 & \pm 1 & \dots & \pm 1 \\ \pm 1 & 1 & \dots & \pm 1 \\ \dots & \dots & \dots & \dots \\ \pm 1 & \pm 1 & \dots & 1 \end{bmatrix}$$

respectively, where $D^{-\frac{1}{2}}$ is a diagonal matrix containing the reciprocals of the saturation coefficients r_{ig}. To get rid of the alternative signs,[3] when they appear, we must (as we shall see in a moment)

[1] Cf. above, p. 149. The original and more general notion of a hierarchy was based, as its name implies, on the broader principle of 'constant order or rank.' On this basis the intercolumnar correlation criterion, which could be derived from the order or rank of the tests in each column, seemed more appropriate than the previous criterion based on proportionality. The distinction is important because, as Dr. Stephenson reminds me, a good many investigators (himself for one) would regard the application of the tetrad-difference criterion as only the most special and stringent form of test (cf. [97], p. 359). Spearman himself has not, I think, explicitly agreed with the identification of a hierarchy with a matrix of unit rank: but it would seem that he would agree that a *perfect* hierarchy might be defined as a matrix having unit rank, except for the diagonal elements, and containing *positive* intercorrelations only.

[2] This is not quite the same as saying that the scores on which those coefficients are based are distributed entirely at random. In that case, it should be remembered, the errors in the correlation coefficients tend themselves to be correlated.

[3] In the usual statement of the intercolumnar criterion, a negative value is taken to imply the presence of more than one factor and the absence therefore of a true hierarchical pattern. Thus, what I term a bipolar hierarchy Stephenson considers should always be analysed into two factors. Thurstone

either reverse the signs of half the columns in the table before correlating or (better still) square the correlations after correlating.

By an 'unadjusted correlation' I mean a product-moment coefficient based on the absolute deviations, instead of on the deviations about the means.[1] Spearman's form of the criterion uses ordinary or 'adjusted' correlations.[2] He himself has drawn attention to two defects attending its use.[3] But in the simpler form in which I have proposed it the criterion would seem to escape both of these objections. In one guise or another, the intercolumnar correlation has been freely employed in the past; and the principle involved has consequently become familiar to students who are unacquainted with the conception of matrix rank. Hence it will perhaps make the simplest starting-point for our discussion.[4]

gives the correct statement: "if the correlational matrix is of rank one then the correlation between any pair of columns is + 1 or − 1." ([84], p. 135: the converse is not necessarily true unless the unadjusted correlation is used.

[1] [102], p. 179 f. Note that (as was there pointed out) "adjusting the absolute product-moment so as to obtain a product-moment about the mean" (though nearly always carried out by those who have used the 'intercolumnar correlation' as a criterion) "spoils the test of proportionality," if by a hierarchy we are to understand a matrix of rank one. Thus, Dr. Carey's example of a perfect hierarchy (*loc. cit. sup.*, p. 2), where the rows of correlations form an arithmetical progression, would *not* be a perfect hierarchy by my criterion, though it would be a perfect hierarchy if the ordinary *adjusted* correlation between columns was employed.

To bring out the analogy with more familiar devices and also to avoid confusing the beginner by introducing more precise technical terms, I shall continue to call the standardized absolute product-moment a 'correlation,' just as I sometimes speak of the absolute root-mean-square as an '(unadjusted) standard deviation.' With a bipolar table, such as a table of residuals obtained with the simple summation method, the mean of each row or column is zero: hence the unadjusted and adjusted correlations and standard deviations are identical. As I have pointed out elsewhere, there are grounds for regarding all correlation tables obtained in psychology as essentially bipolar tables, the ordinary initial table of positive coefficients constituting simply the north-west quarter of a doubly symmetrical bipolar table.

[2] It was first systematically employed by Spearman and Hart in their joint paper on 'General Ability, its Existence and Nature' [24].

[3] *Abilities of Man*, p. ix.

[4] I am much indebted to T. L. Barlow for making a comparative study of the following formulæ (and several others) at my suggestion. In place of my own somewhat elaborate algebraic deduction, based on the 'canonical expansion of the correlation matrix,' I have mainly followed his simplified method of exposition, as being more easily intelligible to the ordinary student. Space compels me to omit the concrete arithmetical examples, which illustrated and gave point to his analogies.

With a covariance matrix, or with a correlation matrix whose self-correlations are known, we have merely to calculate and standardize the product-sums for the rows or columns taken in pairs. To reduce the $\frac{1}{2}n\ (n-1)$ intercolumnar correlations to a single figure, their average has commonly been taken. Let us glance first, therefore, at the criterion which this direct calculation appears to supply.

If in factorizing the original table of observed correlations we follow the method of weighted summation, there is a very simple relation between the factor variances of the correlation matrix and its square or higher powers: we have, in fact (with the usual notation), $R^m = LV^mL'$. With a perfect hierarchy this means that the (unadjusted) product-sum on which the intercolumnar correlation for any two tests is based is simply v_1 times the observed correlation between those tests. Thus, on summing all the product-sums to obtain an average, we have $\Sigma\Sigma p_{ij} = \Sigma r_{ig}^2\ (\Sigma r_{ig})^2 = v_1\Sigma\Sigma r_{ij}$. But now, it may be asked, why correlate a second time? Why not be content with 'first moments'? For, on dividing both sides by v_1, we apparently obtain a very simple test which has indeed been actually employed,[1] viz. $\Sigma\Sigma r_{ij} = (\Sigma r_{ig})^2$: i.e. the sum of the correlations should be equal to the square of the sum of saturations for the one and only 'general' factor. This requirement, it will be noted, is very similar to that which we reached on the basis of the tetrad-difference equation, except that the correlations on the one side of the equation, and the saturation coefficients on the other side, are not squared before they are summed.

To this proposal, however, there are two obvious objections. First, with a bipolar hierarchy, such as that which might be found in a table of residuals obtained by the centroid method, both $\Sigma\Sigma r_{ij}$ and $\Sigma r_{ig} = 0$; hence, if we attempt to sum the residuals (as Thurstone does for his 'first moments' criterion) we are bound to adopt some arbitrary convention for changing negative coefficients to positive. This difficulty, it will be observed, is obviated by the squaring just referred to. Secondly, with the centroid method of calculating r_{ig}, the equation is true of all tables; this particular criterion therefore turns on the identity of r_{ig} as determined by the methods of simple summation and weighted summation respectively. But the simplicity of the reduction thus obtained with the method of weighted summation prompts a practical suggestion: if after all only the sum of the intercolumnar correlations is required, is there not a much speedier way of reaching, or at least of estimating, the final figure?

[1] By J. E. Watson, in an unpublished thesis.

The table of $\frac{1}{2}n(n-1)$ intercolumnar correlations offers much the same problems as the initial table of $\frac{1}{2}n(n-1)$ observed correlations. The number of items now correlated (n) will almost invariably be much smaller than the number of persons correlated for the initial table (N) : only in one or two researches does it exceed 30.[1] Hence, in testing significance, we should in strictness adopt the principles proposed by recent statistical writers for testing small samples. In at least two respects, however, a set of intercolumnar correlations will differ from ordinary correlations : the sampling distribution of the variables correlated, namely, correlation coefficients, is not normal ; and the degrees of freedom will be still further restricted. If we retain this line of approach, it would not be difficult to modify the usual proofs to take both these peculiarities into account. But, as I have indicated below, if we are aiming at great precision, a somewhat different mode of attack would seem desirable. However, with the rough data available in psychology, precise determinations are out of the question. Hence, in an elementary discussion we may ignore these further refinements.

On the other hand, with a chance distribution of the coefficients we may reasonably assume that both the means and the standard deviations of all the columns would be alike. In particular, with a table of residual correlations (for which we chiefly need our criteria) the means, as obtained by the simple summation method (the ' centroid method,' as Thurstone terms it) are zero. It is true, as we have just seen, that the average intercolumnar correlation (even when the correlated coefficients are not distributed at random) will also be zero : for half the correlations will be numerically identical with the reversed half, but will have opposite signs. But the proportionality criterion will remain unaffected if, following the device familiarized by Thurstone and used by him in testing the significance of his residual tables, the variables for one half of the table are reversed in sign.

When the means and standard deviations of each column are identical, there is, as we have already seen, a very simple

[1] Notably in the recent remarkable study by Thurstone to which reference will frequently be made in the sequel [122]. Here 57 tests applied to 240 persons were correlated for the initial correlation table. With the method of correlation employed, however, the standard error is said to be about ·09 (p. 61). Thurstone does not apply his earlier suggestions to demonstrate whether the correlation matrix is of rank one (e.g. the intercolumnar or proportionality criteria [15], pp. 134–5), although, as we shall see, almost all the latent roots except the first are of doubtful significance.

way of approximating to the average intercorrelation, for it is then identical with the intraclass correlation,[1] that is, as we may call it here, with the intracolumnar correlation This, it will be remembered, can be calculated very easily from the ratio of two variances. Moreover, for testing the significance of such a correlation (or of the variation on which such a correlation depends) it is customary to proceed by directly comparing two variances or two standard deviations. We are therefore naturally led to inquire whether, on the hypothesis of non-significance, we cannot adopt this abridged method of computation, and at the same time find some simple ratio analogous to the ratio of two variances, used in the analysis of variance (the '*F*-ratio'), or to the ratio of two standard deviations, used in calculating the more familiar 'correlation ratio' η, for the purposes of the test.

Under the conditions just specified the relation between the average intercolumnar correlation and the corresponding 'correlation ratio' is given by the equation formulated above, namely,

$$\bar{\rho} = \frac{n\eta^2 - 1}{n - 1},$$

where n is the number of variables correlated, and η^2 is ordinarily defined as the ratio of the variance of the means of arrays (σ_m^2) to the total variance (σ_t^2). The test of significance then turns on the ratio—$F = \frac{\sigma_m^2}{\sigma_t^2 - \sigma_m^2}$. When the initial correlation matrix is hierarchical and the intercolumnar correlations are all perfect, η^2 and $\bar{\rho}$ both take their maximum value, namely, 1. But when the initial correlations are distributed by chance, η^2 is not zero, but $\frac{1}{n}$; if, however, we assume that the total variance is n, then $\bar{\rho}$ is zero. Thus, $\bar{\rho}$ may be regarded as a correction

[1] Cf. above, p. 275; also Yule and Kendall [110], pp. 253–8, or Fisher [50], pp. 198–203 (the latter includes an examination of the sampling error of the intraclass correlation).

(or partial[1] correction) of η^2 so that it ranges from 0 to 1. The analogy with the analysis of variance is clear, but for obvious reasons far from perfect.[2] Nevertheless, if, after finding the factor-variances, we base the calculation of η^2 on one or other of the trace-ratios enumerated above, we shall find that several instructive results may be reached.

The Economy Ratio.—First, however, let us glance at the most simple substitution of all. If with Thurstone we could assume that the complete correlation matrix has an exact, assignable rank r (where $r \leqslant n$), we could take the total variance to be analysed, namely σ_t^2, as approximately equal to $r\sigma_m^2$, and write $\eta^2 = \frac{1}{r}$. The above formula for $\bar{\rho}$ would then reduce to an easily intelligible expression, depending essentially on the ratio of the number of tests (n) to the number of common factors (r), or, in more technical language, on the ratio of the order of the correlation matrix to its rank. We should have $\bar{\rho} = \frac{1}{n-1}\left(\frac{n}{r} - 1\right) = E$ (say), an expression which I have called the 'economy ratio.' When $r = n$, $E = 0$; and the factorization leads to no economy in the number of variables. When $r = 1$, $E = 1$; the economy is perfect, and the matrix hierarchical. And between these limits E might be regarded as measuring the simplicity of the correlational pattern, and therefore of the group of processes tested.

With an empirical correlation or covariance matrix, however, to

[1] Strictly $\eta^2 = \dfrac{\sigma_m^2 + \frac{1}{n}\sigma_r^2}{\sigma_t^2 + \frac{n-1}{n(N-1)}\sigma_r^2}$. Thus, if the correlation in the 'infinite population' is zero, η^2 tends to the value $\frac{N-1}{nN-1} = \frac{1}{n}$ when the sample is large enough for us to write N for $(N-1)$ after the fashion of the older formulæ. On the need for correcting η^2 see Kelley [47], p. 240 and references.

[2] In particular, here and subsequently, the usual indications for the degrees of freedom will not hold. For more precise work, however, an entirely different line of attack would seem desirable: e.g. it would be better to deduce the expression for the total variance either from the distribution of the original scores or from independent data showing the amount of unreliability included in those scores. This, however, leads to complicated expressions and laborious calculations, such as the inexact data of the psychologist would hardly warrant.

assign an exact rank is hardly practicable. We know that the majority of the factors have no statistical significance; but, until we have applied a criterion of significance, we cannot say how many of them can be treated as non-existent. We must, therefore, fall back on a less summary method of determining the ratio involved.

According to the nature of the precise characteristic we are proposing to test, we may take the ratio between the observed and the maximum values for the standard deviations (or variances) of (i) the initial correlations or their residuals, (ii) the saturation coefficients derived from these correlations or residuals, (iii) the factor-variances derived from these saturation coefficients, or (iv) yet higher 'moments' derived from the factor-variances by carrying the same process to a further stage.

(i) *The Ratio of Residuals.*—Let us begin by examining the means and variances of the residual correlations, since this form of comparison has been adopted by Thurstone in his last, most interesting research. Consider the matrices of residual correlations, obtained after eliminating $s, s+1, \ldots$ factors from an empirical $n \times n$ correlation matrix. The relative importance of two successive factors may be expressed by comparing either the mean deviations or the standard deviations (or variances) of the two successive sets of residuals. If (as usual) the residual matrices are bipolar, we can take the ratio of the variances to be

$$\frac{\Sigma\Sigma r_s^2{}_{+1}}{\Sigma\Sigma r_s^2} = \frac{v_s^2{}_{+1} + \ldots + v_n^2}{v_s^2{}_{+1} + v_s^2{}_{+2} + \ldots + v_n^2}.$$

Let us first put $s = 0$. We then have, for the first of these ratios, $\frac{\sum_2^n v^2}{v_1^2 + \sum_2^n v^2}$. This ratio will certainly vary with the size of v_1, and will therefore give some indication of the hierarchical tendency of the initial matrix. But as a measure of that tendency it is not cast in so convenient a form as the ratios already proposed.

Let us now put $s =$ number of significant factors. Since Thurstone deals with a reduced correlation matrix of minimum rank, we may designate that minimum rank $(n - t)$. Then the rank of the residual matrix will be (say) $u = n - s - t$. Hence $(s + t)$ of its n latent roots (or factor-variances, as we call them, in dealing with a correlation table) will be zero. But if all the significant factors have already been extracted, then we should expect the u remaining non-zero factor-variances to be approximately equal. Then $\Sigma\Sigma r_s^2 = \sum_{s+1}^{n-1} v^2 = uv^2$. After extracting one more factor, the

sums of the squares of the new residuals will be equal to $(u - 1)\, v^2$. Hence the ratio of the two sums $\frac{\Sigma\Sigma r_{s+1}^2}{\Sigma\Sigma r_s^2} = \frac{u-1}{u} = \frac{n-s-t-1}{n-s-t}$; and the ratio of the standard deviations of the two successive sets of residuals will be $\sqrt{\frac{n-s-t-1}{n-s-t}}$. Accordingly, when the ratio reaches this value, it might be supposed that the residual correlations whose squares have been summed for the denominator are determined entirely by chance.

Thurstone plots a curve for the standard deviations of the first 13 sets of residuals obtained in his research ([17], p. 64); and the foregoing result fits the prolongation of his curve sufficiently well. He offers, however, no criterion based on these standard deviations; but proposes instead an empirical rule based on a similar comparison of sums or means, i.e. upon the mean deviations. When s significant factors have been extracted, " so that only chance variation remains in the residuals,"[1] then, he suggests, $\sqrt{\frac{\Sigma\Sigma r_{s+1}}{\Sigma\Sigma r_s}} = \frac{n-1}{n}$, where $\Sigma\Sigma r_s$ and $\Sigma\Sigma r_{s+1}$ denote the numerical sum of the values of the residuals, calculated regardless of sign. This rule, however, does not make allowance for the fact that we are dealing with ' reduced ' matrices. To indicate the *approach* to equality, the foregoing argument suggests we should rather take

$$\frac{\Sigma\Sigma r_{s+1}}{\Sigma\Sigma r_s} = \frac{u-1}{u} = \frac{n-s-t-1}{n-s-t}.$$

Let us test this formula by applying it to Thurstone's own data. His correlations are based on 57 tests. Consequently the minimum rank of the initial correlation matrix will be 47. After extracting 13 factors, the rank of the residual matrix should be 34. Hence the foregoing formula suggests that the ratio of the sums of the two successive residuals will be approximately $\frac{33}{34} = .971$, provided the remaining factor-variances are equal. Actually, the observed ratio (based on factorization by simple, not weighted, summation) is given by Thurstone as .960, and is apparently still rising: the preceding figures are .921, .945, so that it would seem that equality has not yet been reached.

For Thurstone's two 8-variable tables with no common factors we should have 4 as the minimum rank; the ratio would therefore be $\frac{3}{4} = .750$: Thurstone's calculations give .774 and .742 respectively, averaging .757. For his three 20-variable tables with 1, 2, and 4 significant factors to be extracted, his own criterion gives the

[1] *Loc. cit.*, p. 66.

same figure for all three cases, namely, $\cdot 950^2 = \cdot 903$; mine gives ·923, ·917, and ·900 respectively. It seems clear that a diminishing, not a constant, ratio is required; the observed figures are ·901, ·874, ·847. They are followed (where further figures are given) by higher figures in each case: no doubt, therefore, the selection of 'random coefficients' was not perfectly random.

There is at least one strong objection to Thurstone's statement of the position. If we accept the usual borderline for statistical significance, there will, as a rule, be a wide interval of uncertainty after we have eliminated the s significant factors and before we reach figures in which "only chance variation remains." The transition from significantly diminishing factor-variances to factor-variances that are virtually equal will not be sudden and abrupt. Thus, with Thurstone's data I find it difficult to believe that any factors after the 5th or 6th at the very outside are really significant as judged by the ordinary convention. From that point the factor-variances should become increasingly equal; and at the 47th all the common factors, significant or non-significant, should (on Thurstone's principles) be exhausted, since 47 is the rank of the 'reduced matrix.' Actually, however, the factor-variances are much higher than the probable errors would suggest, and there seems little doubt that, with Thurstone's method of extracting factors, even after 47 had been extracted, there would still be appreciable but not genuinely significant residuals.[1] I doubt, therefore, whether a criterion along these lines is satisfactory and, even if it were, whether it would be applicable to the residuals in question.

(ii) *First Moments.*—In any given table of correlations or in any set of residuals obtained from them, the presence or absence of a conspicuous hierarchical pattern depends, as we have seen, on wide differences between the factor-variances. This means, first, that the first factor-variance must be relatively large; secondly, that the other factor-variances must be relatively small. But, with bipolar factors, if the first factor-variance is to be large, the differences between its factor-saturations must be wide. Accordingly, let us consider this point first of all.

The saturations that describe a factor are obtained, as we have already noted, by a process of averaging. In a sense, therefore, they are weighted class means. Consequently, we may attempt to

[1] I note, in studying figures in research theses, that residuals given after a number of factors have been extracted are often artificially enlarged: the chief causes seem to be using simple summation, constantly inserting the highest correlation in the leading diagonal, and not infrequently rounding off the initial correlations too rapidly.

evaluate the significance of the differences between these saturations along the lines followed in analysing variance. If we start with a table of residuals or with correlations obtained with a homogeneous population, the first factor will be bipolar, the means of the saturations will be approximately zero, and the adjusted variances approximately equal to the unadjusted. If we start with a table of initial correlations that are all positive, we can calculate the adjusted variance of the first factor-saturations in the usual way, or assume that the initial table is really part of a symmetrical bipolar table. For brevity I shall here assume that bipolar conditions may be presupposed throughout. We can then take the variance of saturations as proportional to $\sum_i f_{i1}^2 = v_1$; and the total variance as proportional to $\sum_j \sum_i f_{ij}^2 = v_1 + v_2 + \ldots + v_n$. This latter expression also indicates the maximum value that v_1 can take; it is at the same time identical with the sum of the variances of all the tests—the total variance in fact which our factorization analyses. Let us therefore put

$$\eta_1^2 = \frac{\sum f_{i1}^2}{trR} = \frac{\sum f_{i1}^2}{\sum\sum f_{ij}^2} = \frac{v_1}{v_1 + v_2 + \ldots + v_n} = \frac{v_1}{v_1 + v_r}$$

where v_r denotes the residual variance. We thus reach a more concrete interpretation of the first of the three 'trace-ratios' suggested above. If we regard η_1 as an ordinary correlation ratio, we may test its significance by applying one of the simpler standard formulæ.[1]

We may note that, with this interpretation of η^2, (*a*) for a perfectly hierarchical distribution of the correlations, $v_1 = \Sigma v_f$, $\eta_1^2 = 1$, and $E = 1$ as before; (*b*) for a perfectly random distribution of the correlations, $v_1 = v_2 = \ldots = v_n$, $\eta_1^2 = \frac{1}{n}$, and $E = 0$ as before. With intermediate cases, however, we have no data for directly estimating the total variance, Σv_f. If, as is here assumed, we are adopting a summation method with minimal rank, the figure for Σv_f is merely a theoretical lower limit: the upper limit is n.[2]

[1] Yule and McKendall, *loc. cit.*, pp. 409, 453–4. Fisher, *loc. cit.*, p. 245. But see the comments in the following paragraph, which show that this is only a rough and practical test.

[2] The alternatives are similar to those described by Thurstone, when he distinguishes between working with a 'reduced correlation matrix' R_0 (in which the 'communalities' are substituted for the 'complete test-variance') instead of with the 'complete correlation matrix' R_1 (in which the 'diagonal entries are unity') ([15], p. 66). Where it is necessary to distinguish between the alternative formulæ, I shall affix the same sub-

Hence we can only make a broad estimate of the general region in which factors begin to be non-significant.

Nevertheless, the ratio to which we have thus been led offers the simplest way of describing the relative importance of the several factors. If we take the upper limit, we are virtually assuming that the test-variances are all equal to unity, so that their sum $\Sigma v_f = n$. *The contribution of the* k*th factor to the total variance may then at once be expressed by the fraction* ${}_1\eta_1^2 = \frac{v_k}{n}$. This method was systematically adopted by Miss Davies in comparing the importance of the 'general factor' with that of other factors in a long list of researches on correlating persons [1]; and is also incidentally employed by Hotelling in his illustrative example ([79], p. 434).

If we take the lower limit, with $\Sigma v_f < n$, then it will generally be better to use the 'corrected' formula,

$$\bar{\rho}_1 = \frac{n\eta^2 - 1}{n - 1} \qquad (\eta^2 = {}_0\eta_1^2)$$

$$= \frac{(n-1)\ v_1 - v_r}{(n-1)\ (v_1 + v_r)}$$

This gives a formula analogous to the intracolumnar correlation and a figure analogous to the average intercolumnar correlation. As before $\bar{\rho}_1$, unlike η^2, can take values between 0 (for a perfectly random distribution) and 1 (for a perfect hierarchical distribution). With certain plausible assumptions, we can develop tests along the usual lines for the significance of this expression, which, it will be observed, takes n into account as well as N into account.

(iii) *Second Moments.*—But we also require to test the significance of the differences between the several factor-variances themselves. We may attempt this in two ways. First, we may take the difference between each pair separately; and in particular we may proceed (as Thurstone does) by comparing results from successive residual tables. Then, for a rough approximation we may regard these two tables as depending almost entirely on two independent factors; and we may consider the significance of a single correlation which could be analysed into these two dominant factors. In such a case $n = 2$; and the formula just described reduces to

$$\rho = \frac{v_s - v_{s+1}}{v_s + v_{s+1}}$$

script, and write, for instance, ${}_0\eta_1^2$ and ${}_1\eta_1^2$. (To call the former 'reduced *eta*,' as I have previously done, proves a little misleading, since the 'reduction' of the variance magnifies *eta*.)

[1] *Brit. J. Psychol.*, XXIX, p. 413.

Here $\sqrt{v_s}$ and $\sqrt{v_{s+1}}$ are standard deviations about the principal axes of the concentric frequency ellipses, and are consequently proportional to the lengths of those axes. They thus measure the tendency of the ellipses to broaden into circles or to 'condense' into a single straight line.[1] The above expression may accordingly be described as the 'condensation ratio' for two factors. ρ can be treated as a correlation between two variables in a population-sample of N individuals; and the usual tests of significance applied. This method of determining whether one factor-variance is significantly greater than another has, in fact, been proposed by Hotelling ([79], p. 434). It will serve to show when we have reached a pair of residual factors whose difference is so small that it cannot safely be attributed to anything but chance: it will not serve to show that we have obtained a difference so large that the larger of the two factors is definitely significant.

Systematically and completely carried out, however, this principle would lead us to examine, not merely the $(n - 1)$ differences between successive factor-variances taken in pairs, but the $\frac{1}{2} n (n - 1)$ differences between all possible pairs. But that is precisely the type of situation which the analysis of variance has been devised to meet; and once again we may approach it from the standpoint of an intra-class correlation. Except for the final division by the number of items added, the factor-variances represent the means of the squares of their saturation coefficients; and, just as we have summarized all the differences between the factor-saturations by their standard deviation (or its square), so we can summarize all the differences between the factor-variances by their standard deviation (or its square). Following the same lines as before, we shall be led to a criterion based upon the unadjusted variance of the factor-variances themselves, viz. $\frac{1}{n} \Sigma v_f^2$. The observed value of this expression we can compare with its maximum value; and we obtain another correlation ratio $\eta_{I^2}^2 = \dfrac{\Sigma v_f^2}{(\Sigma v_f)^2} = \dfrac{tr(R^2)}{(trR)^2}$, the second of the three trace-ratios described above. If we compare this formula with the equality reached for 'principal tetrad-difference criterion' (p. 335), we may describe the result as converting the tetrad-*difference* criterion into a tetrad-*ratio* criterion.

If, as before, we put $\Sigma v_f = \Sigma v_i = n$, we can give this ratio a different interpretation. We have ${}_1\eta_{I^2}^2 = \dfrac{\Sigma v_f^2}{(\Sigma v_i)^2} = \dfrac{\Sigma\Sigma f_{ij}^2}{n^2}$. Thus,

[1] Cf. *Marks of Examiners*, p. 255, Yule and Kendall, p. 232, or figure 21 in Brown and Thomson's *Essentials of Mental Measurement*, p. 122.

if the total variance is identified with the number of standardized tests, the tetrad-ratio criterion may be regarded as indicating the variance[1] of *all* the factor-saturations. Once again, it may be noted, the formula gives $\eta_2^2 = 1$ for a perfect hierarchy, and $\eta_2^2 = \frac{1}{n}$ for a perfectly random distribution.

If we require a formula that gives zero for a perfectly random distribution, we may adopt an equation of the same form as before, viz. $\bar{\rho}_2 = \frac{n\eta^2 - 1}{n - 1}$. According to the detailed calculations that Mr. Barlow has been good enough to make, when a chance hypothesis is being tested, $\sqrt{\bar{\rho}_2}$ appears to afford a convenient approximation to the average of the intercolumnar correlations, and so saves calculating those correlations in detail: the agreement is apparently not so close when $\sqrt{\bar{\rho}_2}$ and the intercolumnar correlation are large.

There is, however, a more interesting interpretation to the formula. If we attempt to interpret it geometrically, we see that it is a multi-dimensional version of the formula given above: $\sqrt{\bar{\rho}_2}$ is, in fact, the ratio of the *observed* (adjusted) standard deviation of the factor-variances to the *maximum* (adjusted) standard deviation. It might therefore be described as a 'condensation ratio' for n factors. I may add that, when we have already decided on the number of significant factors contained in the matrix (e.g. by the procedure described on p. 339), it will be better to substitute that number (n' say) for n (the number of correlated tests), and then, in comparing various factor-analyses derived from the same table, it will be sufficient to calculate the numerator of the above ratio only, i.e. to take simply the adjusted standard deviation of the significant factors, viz. $\sigma_v = \sqrt{\frac{\Sigma v_f^2}{n'} - \left(\frac{\Sigma v_f}{n'}\right)^2}$ as the basis of comparison.

(iv) *Fourth Moments.*—We saw at the outset that one obvious difficulty in using the intercolumnar criterion with residual tables arose out of the fact that about half the resulting correlations were generally negative, so that their algebraic total was approximately zero. Instead of simply ignoring or 'reflecting' these negative signs (as in calculating a mean deviation), it is from a theoretical standpoint more satisfactory to eliminate them by squaring (as in calculating a standard deviation) (cf. [8], p. 183). With this modification the figures obtained should all be equal to + 1, without any alternative sign.

Now, as the table of observed coefficients approaches either a

[1] The unadjusted variance, if the saturation coefficients are not bipolar.

perfectly hierarchical or a perfectly random arrangement, all the intercolumnar correlations should become numerically equal. By its mode of calculation each correlation is the ratio between a product-sum and the geometrical mean of two square-sums. But if all these ratios are numerically equal, their squares will be equal; and, in order to find their average, instead of calculating the sum of the squared ratios, we can first square the numerators and denominators and then take the ratio of the sums. We thus obtain :

$$\overline{(\rho^2)} = \frac{1}{n^2}\Sigma\Sigma\,\rho_{ij}^2 = \frac{\Sigma v_f^4}{(\Sigma v_t^2)^2}$$

where v_f and v_t, as before, denote the factor-variances and the test-variances.[1]

This expression is obviously related to the last of our trace-ratios ; and, if the preliminary calculations required by the trace-ratio have already been carried out, that formula may be used instead. With either version, however, the arithmetic is somewhat laborious ; and from the few applications that have been made I am inclined to endorse Mr. Barlow's verdict : " the fourth (moment) criterion is of theoretical interest, but on the whole too elaborate and delicate for ordinary practical use." It would seem to be most useful when the first two latent roots of the initial correlation matrix are nearly equal.

The hierarchical criteria that I have attempted to deduce are descriptive in the first instance of the correlation matrix itself. Calculated as trace-ratios, they are independent of any form of analysis. In expressing them in terms of the factor-variances, however, I have assumed that the factors in question would be those obtained by weighted summation (or some closely equivalent method), i.e. that their variances would be identical with the latent roots of the correlation matrix, or at any rate with a set of figures giving a close approximation to the dominant roots. Thus converted they become descriptive of the derived factorial matrix rather than of the initial correlation matrix. It is but a natural extension of the underlying principle to assume that, when computed as ratios of the variances or standard deviations of the factor-variances, the same formulæ will supply comparable criteria for the factorial pattern

[1] Students who have used the same summation device to calculate Spearman saturation coefficients will be aware that, with the appreciable deviations that occur in non-hierarchical tables, the result of the abridged method of averaging often departs widely from the average obtained in the regular fashion. That occurs here : but my object is, not to deduce the criterion from the intercolumnar correlation, but rather to show the relation between the two principles.

itself, no matter how those factors have been obtained. Consequently, they may be used to compare different factorial analyses of one and the same correlation matrix, and in particular to examine the effect of rotation or other linear transformations.

Summary.—Let me briefly recapitulate this somewhat prolix discussion, and indicate what I take to be the most practicable methods at present available. For testing significance the simplest method is to calculate χ^2 from the squared residuals. This is much speedier than calculating all possible tetrad differences. For estimating hierarchical tendency the best criteria would seem to be those based on first or second moments. If we are interested more particularly in the relative contributions of the *single* factors (e.g. in the importance of the first factor—'*the* general factor') then the simplest procedure is to take the (unadjusted) variance of its saturations, $\eta_1^2 = \frac{v_1}{n}$, except where the number of tests is much greater than the number of significant factors: in that case it will be wiser to examine the contribution to the total *significant* variance instead of the contribution to the total *standardized* variance, i.e. to take ${}_o\eta_1^2 = \frac{v_1}{\Sigma v_f}$ instead of ${}_1\eta_1^2 = \frac{v_1}{n}$. If we desire a single index for *all* factors, then, if n remains unaltered in the tables to be compared (as in comparing factors from the same correlation table before and after rotation), the simplest procedure will be to take the adjusted standard deviation of the two sets of factor-variances. In comparing results from different correlation tables, where n may be different, it will be better to take the tetrad-difference ratio, η_2^2. As before (except where the number of tests is much greater than the number of significant factors) we may assume $\Sigma v_f = n$, and the formula is the equivalent to examining the (unadjusted) variance of all the saturations. The 'condensation-ratio' ($\bar{\rho}_2$) requires slightly more elaborate calculations; but would seem to have both the best theoretical basis and the most intelligible meaning, and so to be most appropriate for formal comparisons on a systematic scale.

Illustrative Results.—The value of these various criteria will be more evident if I give one or two brief instances of their application.

(*a*) *First-moment Criteria.*—First of all, it is instructive to note that, when such criteria are applied to data obtained in psychology either by correlating tests or by correlating persons, the results obtained from different tables evince (with a few explicable exceptions) a somewhat remarkable uniformity. Let us begin with the simplest criterion of all, namely, $_1\eta_1{}^2 = \frac{v_1}{n}$. This is a useful and common device for estimating the importance of any given factor. Its application at once leads to a suggestive result. Provided neither the battery of tests nor the sample of persons is the outcome of specialized selection, and provided n is not so large as to disturb the method of comparison, it appears that the relative contributions of the factors to the total variance diminish in a fairly regular fashion. The general proportions are indicated in the first column of Table VII.[1] When n is exceptionally large and the tests are more heterogeneous than usual, the proportions incline towards the lower of the values shown ; and, when the probable error is high, the steady diminution of the variances begins, as we have already seen, to get arrested as they approach figures of the order of the probable error.

This conclusion would seem to be of some theoretical importance. Writers with such opposite views as Spearman and Thomson have emphasized the apparent fact that nearly all the correlation tables obtained by psychologists show a marked hierarchical character : " the tendency to zero-tetrad differences," we are told, " is very strong in mental measurements " ; " in a complete family of correlation coefficients the rank of the correlation matrix tends towards unity and a random sample from this family will show the same

[1] These proportions are based mainly on an early review of tables of correlations between tests or traits carried out at my suggestion by Miss Jefferson. A similar review for tables of correlations between persons has been more recently undertaken by Miss Davies ; the proportions are much the same, viz. ·48, ·11, ·05, ·03, . . . ([130], p. 412). Similar proportions are obtained with physical measurements, when reasonably comparable.

Taking these proportions and the above criteria, we can reach a rough notion of the size of the sample that would be required to establish 2, 3, or more significant factors by analysing a table of correlations between tests. To obtain evidence that is reasonably conclusive we should require (in round numbers)—for 1 factor at least 20 persons, for 2 factors at least 60, for 3 factors at least 250, and for 4 factors about 1,000 ; but for merely suggestive figures smaller numbers may serve. However, as the fuller criteria clearly imply, the significance of the factors will depend, not only upon the number of persons tested, but also upon the number of tests employed.

TABLE VII

CONTRIBUTIONS OF FACTORS TO TOTAL VARIANCE

Factors.	Theoretical Values.	Thurstone's Data.			
	$\frac{v_i}{n}$	${}_1\eta^2{}_1 = \frac{v_i}{n}$		${}_0\eta^2{}_1 = \frac{v_i}{\Sigma v_i}$	
		Before Rotation.	After Rotation.	Before Rotation.	After Rotation.
I	·30 to ·50	·339	·072	·487	·104
II	·08 to ·15	·082	·061	·118	·087
III	·03 to ·06	·050	·071	·071	·103
IV	·01 to ·03	·035	·084	·050	·121
V	<·01	·033	·048	·047	·069
VI		·027	·050	·040	·072
VII		·025	·046	·035	·066
VIII		·021	·072	·030	·104
IX	<·005	·020	·049	·029	·069
X		·021	·046	·031	·066
XI		·024	·049	·034	·070
XII		·019	·048	·028	·069
Total	·42 to ·75 · +	·696	·696	1·000	1·000

tendency." [1] I have ventured to suggest that a truer description of the facts is not that the correlation tables tend to have a rank of one, i.e. only a single latent root, but that their latent roots, though numerous, are widely separated and exhibit much the same diminishing proportions—one being nearly always three or four times as large as any other and usually accounting, not, indeed, for all the variance, but for at least half the total significant variance.

Given certain specifiable conditions, these proportions, and their mode of diminution, could, I think, be predicted on theoretical grounds.[2] Hence it may be of interest to compare these general values with those actually obtained in the latest and most extensive research in which factor-analysis has been employed—Thurstone's work on *Primary Mental Abilities* [17]. From the saturation co-

[1] [56], p. 139 f.; [87], p. 91; [132], pp. 5, 316. The tendency is *not peculiar to psychological data*, as these writers imply (*v.* pp. 357 and 486).

[2] Since analysis by multiple factors is a process of averaging deviations about preceding averages, the range of the correlations is reduced to rather more than half at each step: the treatment of the first factor as exclusively positive, instead of bipolar, produces the effect of missing a step.

efficients that he has tabulated, the factor-variances can be readily calculated for his 12 factors. The proportionate values are shown in the last four columns of Table VII. Since n is much higher than usual, the values will be exceptionally low if we treat n as the denominator (column 3). For the present comparison, therefore, it will be better to take as the divisor what Thurstone calls " the significant common factor variance," i.e. the total factor variance for the first 12 factors (Σh_j^2 in his notation), namely, 39·6,[1] in short, to use ${}_0\eta_1{}^2$ instead of ${}_1\eta_1{}^2$. It will be seen that, except for certain explicable peculiarities in the tail, the form of the curve is very similar to that already indicated by the theoretical values in the first column.

One or two points call for a passing comment. (i) The increase in the factor-variances at the 10th and 11th factors is evidently due to Thurstone's change of procedure : " at the 10th factor a refinement was introduced . . . that increases the amount of the total variance that is accounted for by each new factor." The anomalous swelling towards the end of the tail is thus at once explained. (ii) With tests applied to 240 persons we should expect coefficients representing a true correlation of zero to be distributed about zero with a standard error of $\pm$ ·06 and a probable error of $\pm$ ·04 : these figures seem to agree with the standard deviations given by Thurstone for his residuals. Accordingly, it is only natural to find the factor-variances diminishing far more slowly when they reach this level. That explains the further peculiarities towards the end of the table. (iii) Since, however the tetrachoric correlation was used, Thurstone tells us that their standard error should be placed at a higher figure than usual, namely, about ·09 (p. 61). Now the standard deviation of the first set of residuals gives a figure of only ·127 and of the second set ·098. After all, therefore, it may be doubted whether more than one or two factors at most can be considered as definitely established, and whether the evidence for more than two or three others is even suggestive. On external grounds I certainly suspect that at least half a dozen factors are probably operative : but the intrinsic evidence by itself can hardly be cited as *disproving* Spearman's doctrine of a single ' general factor.'

(*b*) *Second-moment Criteria.*—But, as I have already argued, the real issue is not to prove or disprove the existence of such a single

[1] Thurstone, of course, does not profess to analyse the total amount of variance (i.e. n, since his tests are presumed to be in standard measure), but only that part of it that is attributable to common factors : this amount in turn is made as small as possible by keeping the number of common factors as small as possible.

'general factor,' but to show how completely the correlations as they stand can be explained by the high variances, not of one factor but of a very few. Let us therefore turn to a measurement of the general hierarchical tendency, and from 'first-moment' criteria to 'second.' As before we may begin by considering the theoretical figures suggested by the general trend of the earlier correlation tables previously published. Taking the higher of the theoretical proportions in Table VII, we may say that the *unadjusted* standard deviation of the factor-variances tends in most tables to rise towards a figure of $\cdot 53\sqrt{n}$. (With an indefinitely large number of variables, therefore, the *adjusted* standard deviation will also tend towards that figure, so long as the correlations remain preponderantly positive: with a limited number of tests, the adjusted standard deviation will be much lower.) If the total variance remained the same, the corresponding standard deviation for a perfect hierarchy would be $\cdot 75\sqrt{n}$. Hence for the ordinary type of table the tetrad-difference ratio $_0\eta_2^2 = \frac{\Sigma v_f^2}{(\Sigma v_i)^2}$ will be in the neighbourhood of $\cdot 50$; for a perfect hierarchy the figure would, as we have seen, soar to 1·00. We may thus say that, as a general rule, a ratio appreciably above $\cdot 50$ may be considered as indicating an unusually close approach to the hierarchical pattern; and a ratio appreciably below $\cdot 25$ an unusually close approach to equal or equalized factor-variances.

Let us compare this with actual results.[1] I shall again choose the two tables analysed by Thurstone, chiefly because his conclusions seem at first sight most strongly opposed to Spearman's views on hierarchical tendencies. At the same time his figures will enable us to see how far the theoretical criteria are applicable to data which fall short of the requisite conditions in two or three common respects: first, an approximate method of analysis—simple summation instead of weighted—has been employed, so that the factor-saturations are not strictly independent, and the factor-variances are not exactly equal to the latent roots; secondly, not all the factors have been extracted; thirdly, the high probable errors have apparently increased the variances of the less dominant factors.

Let us begin with Thurstone's illustrative analysis of Brigham's table for 15 intelligence-tests ([15], pp. 108 *et seq.*): this is typical of the more usual table, and Thurstone has also appended an analysis by weighted summation. With the latter method (p. 133)

[1] Once again I am indebted to calculations made by Mr. Eysenck (for Thurstone's work) and Mr. Barlow (for that of earlier investigators).

the factor-variances are approximately $\cdot 34n$, $\cdot 09n$, $\cdot 03n$, and $\cdot 02n$. The unadjusted standard deviation is $\cdot 34\sqrt{n}$; the total communality is 6·965, i.e. $1 \cdot 80\sqrt{n}$; and the tetrad-difference ratio η_{12}^2 is ·554. With the method of simple summation (pp. 117, 131) the factor-variances for the first three factors are slightly diminished and that for the last slightly increased; this reduces η_{12}^2 to ·552. The difference is negligible. And thus, whichever way it is analysed, the correlation table shows a hierarchical tendency quite as marked as any that would be found in the majority of earlier tables, including those of the Spearman school.

When we turn to the figures obtained from *Primary Mental Abilities* ([17], pp. 113–4), we obtain much the same figure for the unadjusted standard deviation, namely, $\cdot 36\sqrt{n}$; but Thurstone's figure for the total communality is now $39 \cdot 62 = 5 \cdot 25\sqrt{n}$. Owing to the increase in the community the figure for the tetrad-difference ratio is much lower, namely, $\eta_{12}^2 = \cdot 27$. The reduction is due to the extraction of an exceptionally large number of factors (12 in all) of which the majority make (as will be seen from Table VII above) only small and probably non-significant contributions—though even these seem to have been exaggerated by the method of calculation (particularly, I imagine, by the high probable error introduced by the tetrachoric procedure). These more doubtful factors tend rather to obscure the effect of the few dominant and significant factors.

Effect of Rotation.—The comparisons are simpler when we turn to study different factorial matrices obtained from the same correlation table, and therefore based on the same number of tests and, as a rule, on the same number of factors. For this problem we decided, it will be remembered, to work with the adjusted standard deviations or variances, and calculate a 'condensation-ratio.'

With Brigham's correlation table, the number of factors actually extracted is very small. We shall therefore assume that the variances of the unextracted factors are approximately zero. For the adjusted standard deviation of the factor-variances we then obtain a figure of 1·26 with both simple and weighted summation; the maximum would be 1·74. The condensation-ratio (ratio of the observed and maximum values of the adjusted variances) is thus $\rho_2 = \cdot 52$ (the figure could, of course, be calculated directly from the value for η_{12} by the formula given on p. 353). Now, when Thurstone rotates the axes to obtain a 'simple structure,' the new factor-variances are—verbal factor 2·06, numerical factor 0·91, and visuo-kinæsthetic 0·72 (p. 169); the standard deviation of these and the remaining factor-variances must therefore be well below 0·57.

Thus the rotation reduces the condensation ratio to $\rho_2 = \cdot 10$ or less. This means that, instead of concentrating or condensing nearly all the variance into one dominant factor, the demand for 'simple structure' has spread the variances more equally among the three 'group-factors.'

When the number of tests and factors is large, the simplest procedure will be to base the standard deviation on the significant factors only. In *Primary Mental Abilities* Thurstone extracts 12 factors for his 57 tests. If we assumed that there were $(57 - 12) = 45$ more factors, all with a variance of zero, we should be misrepresenting his procedure, since the centroid method claims to be based in theory on a matrix of minimum rank and in practice on the smallest number of significant factors. For the 12 factors actually extracted we find that the standard deviations of the variances are 4·93 before rotation and 0·74 after rotation. Once again, therefore, the rotation produces a most remarkable flattening. Such effects, as we saw at the outset, were to be anticipated on general grounds; and in this instance the result was obvious at a glance once a parallel table of factor-variances had been compiled. But by using a quantitative criterion, we are now able to compare the amount of such changes in different cases.

One or two of the 'rotated factors,' it will be seen, stand out as more important than the rest, notably those which Thurstone terms *V* ('Verbal relations'), *R* ('Reasoning'—a miscellaneous group that Thurstone finds it difficult to classify), and *S* ('Visuo-spatial,' apparently corresponding to the 'perceptual *g*' of Spearman and Stephenson). Now the cognitive processes thus specified are precisely those which are held to characterize the most efficient tests of general intelligence. Once again, therefore, it is evident that the results are not so inconsistent with the theory of general intelligence as Thurstone suggests. Spearman himself would emphasize the common characteristics shared by all these tests—namely, the eduction of relations and correlates: Thurstone, on the other hand, stresses the specific or material characteristics peculiar to each one. That these latter are relatively inessential to the definition of the factor seems obvious when we discover that among the tests which have high saturations for the 'visuo-spatial' factor are 'Syllogisms,' 'Verbal Classification,' and 'Sound Grouping' (classification of words according to similarities of sound)—tests which are neither visual nor spatial but must depend largely on the eduction of logical relations.

Such figures do not necessarily invalidate the results given in the factorial matrix obtained by rotation. On the contrary, as we have

already seen, those results are very much what we should expect with discontinuous groups of tests. But in that case it would appear more natural to seek such a structure from the outset, and employ the simple group-factor method.[1]

In conclusion, let me deprecate the idea that a ' simple structure ' of the kind illustrated by Table I on p. 151 of Thurstone's book is necessarily to be sought in any and every analysis, and that its

[1] In this case it seems unnecessary to proceed with the more elaborate calculations and compare the condensation-ratios. It may, however, be of interest to append Mr. Eysenck's further computations. Taking Thurstone's figure for the total test-variance, namely, 39·62, he at once obtains for the " maximum variance, unadjusted and unaveraged," i.e. $(\Sigma v_t)^2$, a figure of 1569·74. The actual variance, $\Sigma v_f{}^2$, is before rotation, 421·80, after rotation 138·00. Hence the tetrad-difference ratios, $_0\eta_2{}^2$, are ·27 and ·09 respectively and ρ_2, ·22 and ·03 respectively. Here we notice that the retention of a larger number of factors with low variances in the unrotated matrix already produces a somewhat unusual degree of spreading. He also shows that the group-factor method yields, by a very quick and easy procedure, a ' simple structure ' very similar to Thurstone's own, supplemented by a general factor of the Spearman type which accounts for the peculiarities described in the following paragraph in the text. (I am much indebted to Mr. Eysenck for allowing me to incorporate some of his calculations and comments in the above discussion ; a fuller account of his work will be found in his doctorate thesis).

I may add that, if one ' general ' factor is retained, and the other common factors are replaced by group factors, the flattening will, of course, be less marked, but still, as a rule, quite conspicuous : thus, with my own data for scholastic tests, the four significant factors extracted by weighted summation have a standard deviation of 1·71 ; when they have been rotated to one general and three group-factors, their standard deviation is reduced to 1·19. The effect of rotation, it will be seen, is roughly to halve the variance ([128] p. 55).

More recently still, another of my former students, Dr. J. G. Taylor, has subjected one of Webb's correlation tables for character qualities to a rotation somewhat resembling that advocated by Thurstone : on calculating the factor-variances from his saturation coefficients, I find that their standard deviation is 1·79 before rotation, and 0·68 after rotation. Here the variance is reduced to one-seventh ; and once again the transformation to a ' simple structure ' tends to level out the differences between the factors (*Brit. J. Psychol.*, XXX, pp. 158, 161). If, however, we retain a general emotional factor, as Webb himself and I should be inclined to do, then, with the group-factor method, a factor-pattern is quickly obtained which fits the initial correlations more closely, and is not only consistent with the conclusions already drawn by Webb and Maxwell Garnett from these data, but also exhibits the more specialized character-qualities on which Dr. Taylor lays chief stress.

validity is sufficiently guaranteed by the elegance of the mathematical demonstration. That, indeed, is not my own interpretation of Thurstone's statements: but other readers seem to have inferred some such intention. May I, therefore, repeat that, in my view, the kind of structure to be required in a factorial matrix is not something that can be laid down *a priori* once for all, but something that is necessarily determined by the selection of tests in each particular case? In every psychological field it would be quite easy to choose a set of tests which almost certainly would, and another set which almost certainly would not, conform with a pattern like that illustrated in the table cited.

Thus, Brigham's 15 tests—6 definitely verbal, 4 definitely numerical, and 5 definitely visuo-kinæsthetic—naturally lead (as we have seen) to group-factors that do not overlap—a 'simple structure' of the step-ladder type. But we might easily include tests that overlapped in cyclic fashion, or in several varying directions at once. The other set of tests that we have examined here, the battery used in *Primary Mental Abilities*, shows this more irregular constitution. How far these further complications will assist or hinder factorial deductions I need not here discuss. My own opinion is that, in planning experiments beforehand, we should select our tests, so far as possible, in such a way that absence of overlap will be frequent enough to provide the necessary data for distinguishing between the factors shared by wider groups and factors common to narrower groups only. This follows from the principles on which the group-factor method is based.

Even the size of the factor-variances is very much at the mercy of selection. When dealing with discontinuous groups of tests, for instance, those in Brigham's study, we could, by reducing (say) the number of verbal tests from 6 to 4 or 3, and by increasing the number of numerical tests from 4 to 6 or 7, greatly alter the relative proportions of the factor-variances. To insist *a priori* that every factor must have at least one zero, i.e. that there shall never be any factor common to all the tests, is not only an unfair predetermination of an empirical issue, but also a demand which (when we remember that factors in the actual performances must include not only abilities but also the conditions of testing and correlated errors of measurement or observation) is highly unlikely to be fulfilled.

CHAPTER XV

SUMMARY AND CONCLUSIONS

THE main conclusions we have reached may be summarized as follows:

1. The various procedures put forward by different writers for analysing a matrix of test-scores or mental measurements may be classified according to a simple scheme, and appear to be related to one another by simple algebraic relations.

2. Of the two chief lines of approach the *analysis of variance* has much the same objects as the factor-analysis of correlations, and may advantageously be used to solve many of the problems hitherto attacked almost exclusively by the latter. The factor-measurements for the general factor, obtained in analysing correlations between tests, are essentially means of 'classes' or 'arrays,' whose variance is tested for significance in the analysis of variance. If the test-measurements are in standard measure and simple summation is used, the average saturation of the general factor is virtually equivalent to the ratio of the observed standard deviation of the class means to the maximum total standard deviation.

3. For factor-analysis in the narrower sense the chief procedures hitherto proposed may be classified as derivatives of three main principles: (i) the *group-factor* method; (ii) the *simple summation* method; and (iii) the *weighted summation* method or method of least squares—both the latter producing general factors only. Each has its appropriate uses; and the results, so far from being incompatible (as their authors have alleged), prove to be either linear transformations of, or approximations to, one and the same set of values.

4. Each method may be applied to tables of *covariances* as well as to tables of correlations, and to correla-

tions (or covariances) between *persons* as well as between traits.

(*a*) With the method of weighted summation (least-squares method), when the averages for persons and for tests have been equalized (a process which is virtually equivalent to eliminating the corresponding general factor), the factor-loadings obtained by covariating *persons* are identical with the factor-measurements for persons obtained by covariating *tests* multiplied by the standard deviation of the corresponding factor.

(*b*) The factor-loadings obtained from a table of *covariances*, when divided by the standard deviations of the tests, yield saturation coefficients that will satisfy the corresponding table of *correlations*.

(*c*) The *variance* of mental traits when directly measurable appears to vary closely with their complexity; and, when not directly measurable, may be treated as equal to the complete communality, i.e. to the sum of the squares of the saturations for all factors.

5. The group-factors resulting from the *group-factor method* tend to have positive or zero saturations only: the factors resulting from the *general-factor methods* (other than the first factor) have both positive and negative saturations, and are thus bipolar.

(*a*) The relations between the two sets of *saturation coefficients* can be expressed by a triangular rotation matrix, containing positive and negative multipliers appropriately placed. Such a transformation matrix provides the best means of converting one set of factor-saturations into the other.

(*b*) The relations between the two sets of *factor-measurements* are expressed by the same triangular matrix. This shows that the general-factor measurements may be regarded as derived from the group-factor measurements by taking weighted differences between the latter. Thus the bipolar factors extracted by the general-factor method are essentially 'difference-factors,' and as such may be intelligibly interpreted as they stand.

(*c*) The group-factor method (with overlapping factors if necessary) yields a speedier and more precise method of

obtaining ' simple structure ' than the double process of first analysing the correlation matrix by a general-factor method (e.g. the centroid method) and then rotating the axes graphically.

6. The general-factor methods, whether proceeding by *simple summation* (as in Thurstone's centroid method) or by *weighted summation* (as in Hotelling's method of principal components), are in effect merely alternative ways of approaching one and the same set of values.

(*a*) The *saturation coefficients* and the factor-variances obtained with either method prove to be approximate determinations, more or less exact, of the ideal values for the latent vectors and the latent roots that would be reached by solving the characteristic equation directly. The two sets of figures arise automatically in the course of a converging sequence of successive approximations, those of the simple summation method appearing as the first approximation of all, and those of the least-squares method appearing towards the end of the convergence.

(*b*) The *factor-measurements* obtained by the two methods evince a similar relation. For the first or ' general ' factor the estimates implied by the simple summation method are virtually the mere unweighted averages of the original standardized test-measurements ; those for the next factor represent the averaged deviations about the first ; and so on. The estimates provided by the weighted-summation method are based on the best weighted averages of the original test-measurements, the principle of least squares having been applied at the initial stage, in deducing the saturation coefficients, instead of at the final stage in deducing the regression coefficients. Thus, for any specified number of factors, less than the rank of the initial matrix, the ' least-squares method ' yields the best theoretical fit both to the original correlations and to the original measurements.

7. Various methods have been put forward for determining *the number of significant factors* that can be extracted in any given case, i.e. for determining the significance of the factors successively extracted from a given sample of tests and persons. None of the methods in

common use is entirely satisfactory. The method here proposed turns on the fact that χ^2, as a 'measure of discrepancy,' may be used to test the distribution, not only of frequencies, but of the sums of squares of any sets of variates presumed to be normally distributed with unit standard deviation. Accordingly, the residuals (calculated after making Fisher's z-transformation when necessary) are squared, summed, and the sum expressed as a ratio of their expected variance, $1/(N - 3)$. The probability that this total is the mere effect of sampling errors is then ascertained in the usual way.[1]

8. A related problem is to prove or disprove the hierarchical character of a given correlation matrix. For this purpose, instead of calculating intercolumnar correlations or tetrad differences in detail, it seems simpler and more satisfactory to compute a single index to measure its *hierarchical tendency*. One or other of the 'trace-ratios' may be used for this purpose. These are virtually equivalent to calculating the variances of the factor-saturations and of the factor-variances respectively. The former indicate the contributions of each factor to the total variance; the latter, or a simple function of the latter (the condensation-ratio) may be used to indicate the amount of flattening produced by rotating axes. Estimates of the significance of the factors may also be based on these trace-ratios.[2]

[1] The use of χ^2 is warranted only when the estimates of the parameters employed are 'efficient' ([110], p. 428), i.e. based on the 'method of maximum likelihood' or an approximation thereto ([50], p. 15). The method of least squares generally yields such estimates ([50], p. 23). The centroid or simple summation methods do not.

[2] I should like to express my indebtedness to Dr. Trubridge for his kindness in reading the manuscript of this paper and pointing out several obscurities.

PART III

THE DISTRIBUTION OF TEMPERAMENTAL TYPES

CHAPTER XVI

A REPLY TO CRITICISMS OF THE METHOD

Problem.—The inquiry that follows has grown out of an attempt to analyse part of the large collection of psychological data accumulated during my work for the London County Council. It is included here because it provides in concrete form a practical illustration of some of the more controversial problems dealt with in the preceding pages and of the way I have proposed to meet them. The question with which it is primarily concerned is the distribution of temperamental traits. Are temperamental characteristics, like intellectual characteristics—general intelligence, for example—distributed in accordance with the normal curve? Or does their distribution indicate a number of independent groups or temperamental types?

In order to assess each individual's approximation to the type to which he apparently belongs, I have used a simplified method of factor-analysis, involving the correlation of persons. With the aid of the figures so procured, an ideal or hypothetical graph, as it were, is empirically constructed for the temperamental type, similar to the 'psychographs' I have published for intellectual types; analogous graphs are obtained for the temperamental characteristics of each individual: and, since each graph represents a set of numerical measurements for the same series of traits, the approximation of the individual's contour to the theoretical contour can then be measured by a coefficient of correlation.[1] The method will thus serve to illustrate by a

[1] The use of such psychographs as a "means of recording or classifying individuals as members of a given type" was described in an early L.C.C. *Report* ([35], p. 64, and Figs. IX, 1–4), and has been found exceedingly useful in work on educational and vocational guidance. The employment of "'standard personalities'—typical individuals, that is, who are made to serve as common and constant points of reference" (*The Young Delinquent*,

definite example the particular view of factors and of factor-analysis, which I have endeavoured to advance in the preceding pages: namely, that factors are essentially principles of classification, and that factor-analysis is merely a device for assigning an individual member, whether trait or personality, to its appropriate class on the basis of an average (or weighted sum) of a relevant set of assessments.

The procedure is simple enough in actual practice: but the theoretical assumptions on which it rests have recently become the subject of much criticism. Hence the first two chapters will be concerned more with the discussion of methods than with the demonstration of results. My critics have rightly pointed out that, in my justification of the general principles employed, 'several statistical difficulties are involved which [my earlier exposition] passed over rather lightly in order to present the main idea as clearly and briefly as possible' ([132], p. 286). Students working in the same specialized field have also raised a number of incidental questions which could hardly have been discussed at length in a journal for the general reader. These questions and these difficulties are themselves by no means devoid of theoretical interest; and I have therefore thought it necessary to preface the summary of concrete results by a more detailed defence of the statistical procedure.

My plan will be to take in turn the chief objections urged against the proposals, and to indicate the ways in which they can be met. This will entail a criticism of my critics. But I should like to say at the outset that my object is not to engage in controversy, but simply to use this dialectical procedure as the clearest means of extracting the elements of truth that are no doubt common to both the antithetical approaches.

p. 418)—has proved particularly helpful in regularizing assessments for complex temperamental characteristics, where more objective norms, such as are provided on the cognitive side by standardized tests, are no longer available. The relation between what I called 'temperamental profiles' and the constants extracted by factorizing correlations between persons is briefly indicated in [101], p. 65. The full method of calculation is explained in my paper on *The Analysis of Temperament* [114].

As is inevitable in a statistical survey, I have laid myself under a deep obligation to the many persons who have co-operated. I am indebted first of all to the London County Council for permission to make use of material gathered while I was still in their service. To the numerous teachers, social workers, and research students, who have assisted me in collecting or in analysing the data, I owe a heavy debt of thanks. For help with the arithmetical calculations I am especially grateful to Mr. A. F. Roberts and Miss L. J. Carter, to my secretary Miss G. Bruce, and to Dr. A. J. Marshall, Research Assistant in the psychological laboratory.

Most of all, perhaps, I am indebted to my former colleague, Dr. W. Stephenson, whose questions and criticisms are representative of those put forward by many other readers, and at the same time possess the additional advantage of being based on an unusually wide experience in similar fields of research. A frank expression of disagreement is always more fruitful than the mere expression of assent; and Dr. Stephenson has been good enough, not only to spend many hours in discussing these problems, but also to set down his views on paper in the form of detailed letters and memoranda which could be studied and discussed at length. In another publication [138] we have endeavoured jointly to summarize the several points on which we agree and disagree. Here it will be my purpose to deal chiefly with outstanding points of disagreement there disclosed (*loc. cit.*, pp. 274–80), and, so far as I am able, to defend more fully the position I there briefly outlined.

This, too, is an appropriate place to express my thanks to Professor Godfrey Thomson for his sympathetic criticisms of my method in his invaluable book on *The Factorial Analysis of Human Ability*. He has since been generous enough to suggest that the differences between us may possibly be due, not so much to an actual disagreement over fundamental principles, but rather to the gaps that were almost inevitable in a brief preliminary description. In the following pages the more glaring of these gaps have, I hope, been filled in. Actually, the text of this paper was completed and typed before his book appeared; and consequently a few of his points may still remain unanswered. However, in the first of the three papers in this volume (which was written last of all) I have taken the opportunity to discuss what seem the more important and more general of his criticisms (e.g. those relating to negative saturations or loadings, the assumption of objective differences in the variances, and the use of unstandardized scores). Hence there is less need to take up these more theoretical questions here.

On the concrete issue—the distribution of temperamental types—

I think it will be found that the evidence is reasonably conclusive. But on the wider methodological issues far more research is needed; and here my purpose will be mainly to answer specific criticisms and inquiries, not to attempt any final or systematic survey. Lastly, may I emphasize that a statistical study by itself can but touch the very fringe of a complex psychological problem which in the near future will doubtless be attacked by more direct methods of approach?

Procedure.—In a recent paper on the 'Analysis of Temperament' [114] I ventured to offer further evidence for my earlier classification of temperamental types on the basis of what are commonly termed 'group-factors,'[1] and attempted incidentally to show how the factorial methods adopted in the study of intellectual disabilities might be used for the analysis of abnormalities of character. The emotional characteristics of 124 children, referred to me on the ground of delinquency or nervous disorder and made the subject of detailed observation and report, were submitted to a statistical analysis to determine how far their behaviour could be expressed in terms of a few common underlying tendencies. Following the procedure applied in an earlier research,[2] the assessments for the several traits were first correlated and the covariances and correlations then factorized by the methods regularly employed in dealing with intellectual characteristics. Apart from the well-established factor of 'general emotionality,' which had been deliberately excluded (so far as possible) by working with a specially selected group of cases, three further factors were detected; but of these by far the most conspicuous was a bipolar factor making for aggressive or extraverted behaviour when positive and for inhibited or introverted behaviour when negative.

There is, however, as I pointed out, a second way in which an analysis might be attempted. The classification of persons by temperamental or mental type implies that certain individuals can be assigned to one group, and other

[1] *The Young Delinquent*, 1925, pp. 514–5, and Table XVIII.

[2] *Brit. Assoc. Ann. Report*, 1915, pp. 694–6, 'General and Specific Factors Underlying the Primary Emotions.'

individuals to another group, on the basis of their resemblances and differences in a number of specifiable traits. Unfortunately, among human beings (as contrasted, for example, with the fruit-flies and the primroses which have been so successfully employed for the study of genetic types) the relevant resemblances and differences are far from perfect or complete. Nevertheless their amount may readily be measured by correlating the measurements obtained from one person with those obtained from others.

As we have seen in the earlier pages of this book, the device of correlating persons raises many obscure statistical issues; and the difficulties are not so easily disposed of as in the case of correlations between traits. But there can be little doubt about the practical utility of the procedure. In clinical work, a psychologist has primarily to analyse the composition of the concrete individual, not the composition of abstract abilities or traits; he has to take into account qualitative characteristics for which there are no standardized tests, quite as much as quantitative characteristics, like intelligence or educational attainments, which can be readily measured; almost inevitably, therefore, he comes to think in terms of the perceptible resemblances between his individual cases rather than in terms of supposed affinities between hypothetical functions or propensities. In an academic research on educational abilities, the investigator can ask a teacher to rank or grade the pupils in a single class in order of merit for the six or seven main subjects of the school curriculum; and he can then proceed to correlate school subjects. But in clinical work on backwardness, delinquency, vocational guidance, and the like, each fresh case is reported on by a fresh teacher; and, as we have already noted, it is much safer to compare judgements by the same teacher on the different qualities of the same child than to compare judgements by different teachers on the same qualities of different children. Consequently, when the practical psychologist comes to review his material, he is naturally tempted to begin by arranging his data, not according to the tests, the school subjects, or the traits assessed, but according to the several individuals.

Yet, although he has thus altered the customary mode of

approach, he is not hoping to discover another set of factors, but merely to arrive at the same result by a more serviceable route. And, in proposing the alternative procedure, I expressly made what seemed to me the natural assumption that the factors reached by classifying the persons according to their traits would be virtually the same as those reached by classifying the traits according to the differences shown by the persons. But such an assumption should be susceptible of formal proof or at least of verification when challenged. Accordingly, my previous article went on to factorize the covariances and correlations between the several persons; and I sought to prove, by a concrete arithmetical illustration,[1] that the main resemblances and differences between them could be explained by precisely the same 'bipolar factors' as were discovered when the more familiar procedure of correlating traits had been adopted.

To exhibit the *exact* equivalence between these two modes of approach, somewhat stringent conditions had to be imposed: first, that the group of persons to be compared should be as homogeneous as possible in regard to all relevant qualities except those producing the temperamental types; and, secondly, that the mathematical analysis employed should be one which would give the best possible fit to the empirical measurements actually obtained. The first condition was secured by making up a batch in which the average or total measurement of every person was identical; the second, by applying the method of least squares—a device well recognized in almost every branch of science.

Those who have been good enough to comment on my arguments have rightly urged that it is scarcely practicable for these two conditions to be rigidly observed in ordinary inquiries. Nor are they altogether convinced that what holds good within a group of nervous or delinquent children will be equally true of adults or of the child population as a whole. In what follows, therefore, my first object will be to demonstrate that the same results are obtained even under rougher conditions of clinical investigation and with the simplest methods of statistical analysis. I shall

[1] A general algebraic demonstration had already been offered in [101].

then endeavour to show that, even within the normal population, similar types may be readily detected, although in a random or unselected group the more thoroughgoing specimens will be relatively few.

This will lead to the all-important problem of the distribution of temperamental types. I myself have hitherto maintained[1] that temperamental qualities, like intellectual qualities, are distributed in broad conformity with the normal curve, that is to say, the mixed or intermediate types are the commonest, and well-marked examples of the opposite extremes are about equal in number and equally rare. From this it would follow that (phenotypically as distinct from genotypically[2]) there can be no such things as mutually exclusive temperamental types, but only tendencies towards this extreme or that. Resemblance to the idealized type thus proves to be essentially a matter of degree; and it becomes necessary to devise some practicable means of measuring the degree. Both the methods I proposed, however, and the inferences deduced have been called in question. Hence fuller evidence is much to be desired. Accordingly, from the numerous case-records collected during many years' work for the London County Council, I have endeavoured to extract numerical data which, I hope, may yield a more convincing answer to these various questions.

Criticisms.—The double method described in my previous article was first employed for studying such problems as that of tempera-

[1] E.g. 'Mental Differences between Individuals' (*Brit. Ass. Report*, Presidential Address to Section J, 1923, pp. 215–39 *et seq.*). The opposite view is taken in [138], pp. 273, 279, esp. § 17.

[2] I repeat this reservation more explicitly, because at least one friendly critic has accused me of self-contradiction, since, in discussing the inheritance of temperamental qualities, I suggested that the distinction between the aggressive or sthenic type, on the one hand, and the repressed or asthenic, on the other (and between the physical characteristics which sometimes accompany them), might at bottom be *partly* due to a pair of alternative genetic factors. In reply I should argue that this speculation about discrete *causes* is no more in conflict with the hypothesis that *observable traits* are distributed continuously and normally than is the suggestion that human stature is partly controlled by unit-factors obeying Mendelian laws, although its distribution is approximately normal.

mental and æsthetic types in co-operation with several research-workers at the Institute of Industrial Psychology and at the London Day Training College: our conclusions were that the *first* or 'general' factor for persons (i.e. roughly speaking, the factor that is responsible for the differences between those persons' averages) could not be reached by correlating tests; but that each factor *except the first* (i.e. the 'group-factors' or 'type-factors') remained much the same for all practical purposes, whether it was obtained by correlating persons or by correlating tests (cf. above, p. 175). On resuming the work at University College, I found my colleague, Dr. Stephenson, strongly inclined to doubt—and, I think, quite rightly—whether these inferences had been adequately proved. As a result of his own experiments with the procedure, he was later led to suggest several interesting modifications both in the methods to be used and in the inferences to be drawn.[1] His views in this connexion are deserving of special attention, not only because they are based on formal statistical analysis, but because he is one of the few statistical psychologists with a wide experience of mental testing among patients in mental hospitals. With his conclusions I shall deal later on. Here I propose first to consider his criticisms in regard to method; then, accepting his suggestions as completely as I can, to show how his own procedure leads in the end to virtually the same results as mine.

His examination of my article is, I understand, appearing in a forthcoming number of the same *Journal*; but he has been generous enough to send me in advance a copy of his comments.[2] Similar questions have been put by other writers—psychologists, psychiatrists, and statisticians; and, if I deal more fully with the difficulties advanced by Stephenson, it is mainly because, thanks to the frequent opportunities for discussion, it has been possible to narrow the points of difference or misunderstanding down to fairly definite issues. Most of them, though not perhaps all, arise from the incompleteness with which, for the sake of brevity, I was forced to describe the statistical procedure employed and the arguments on which that procedure was based.

The chief objections may be summarized under four heads.

(1) Stephenson's strongest criticisms are directed against my use of covariances instead of correlations. "Professor Burt," he writes,

[1] The modifications have been briefly described above, pp. 182–8.

[2] A brief statement of his main criticisms is to be found in the postscript added to his article in *Psychometrika* (III, pp. 206–9), which deals with the more general proof given in [101]; a summary of the more specific points is contained in his section of our joint article [138]: cf. also [136].

" has put forward the claims of covariance analysis. He works with unstandardized scores and covariances. Unfortunately, in psychology, there are no generally accepted units (other than the standard deviation). In our intelligence tests, for example, we can obviously make the raw scores and variances what we like : both are purely arbitrary matters, dependent upon the mere whim of the psychologist. Clearly any principle involving such arbitrary foundations can scarcely merit a moment's serious consideration." This is a criticism which I have already endeavoured to meet (cf. above, pp. 282–6). As regards traits, I readily agree that in psychology the variances may be, and perhaps very commonly are, " purely arbitrary matters " ; and I should equally maintain that, from the statistician's point of view, this is ' unfortunate.' But I do not admit that it is entirely unavoidable. Stephenson cites the analogy of intellectual tests. But here it is by no means difficult to show that the amount of variance must differ appreciably from one intellectual process to another, since it evidently depends on the complexity of the process and also (it would seem) on the degree to which it has become automatic (cf. pp. 283 f.).

When we are working with temperamental assessments, the matter is more obscure. Yet even here I should have thought that no one (except for convenience of calculation) could really suppose that the variance for every trait or every person was precisely the same. In describing the particular data with which my article was concerned, I gave special reasons why it seemed unnecessary, and indeed undesirable, to treat the variances and the standard deviations as equal for all the traits. The exceptionally wide range of the marks for ' fear ' and ' anger,' for instance, was due, not to the accidents of the psychologist's tests, but to the fact that exceptionally wide differences of observed behaviour had been actually recorded in the different persons studied. We had, in fact, picked out a number of fearless delinquents and over-anxious neurotics ; and it would seem very difficult to assume that the range of behaviour and the variance of the measurements were the same for ' fear ' and ' anger ' as they were for (say) ' curiosity.' Since, however, it was also of interest to consider what might happen in a ' selected standard population ' (i.e. a group in which the individual variation might conceivably be regarded as approximately equal for every trait, so that the standard deviations could be set at unity throughout), I went on to show that the factors derived by my method would fit the correlations quite as well as the covariances.

When we turn from traits to persons, it is often difficult to discover what precise meaning is to be attached either to ' standard-

ization' or to 'correlation': both now seem far more at the mercy of "the psychologist's whims." Stephenson postulates that the distribution of trait-measurements within one and the same person must fit the normal curve and that its standard deviation should be the same for different persons ([138], p. 279 and refs.). This would be obviously incorrect for intellectual traits, such as are ordinarily measured by mental or scholastic tests: one child's measurements may be nearly the same for all the traits tested, and another may be extremely gifted in one trait and extremely lacking in another (cf. [35], Figs. IX, 1 and 4). As regards temperamental traits, the list compiled by Stephenson himself can scarcely be supposed to fit a strictly normal distribution or to show equal variability from one person to another. However, for a preliminary classification of persons into groups or types, it is sufficient to judge by broad resemblances; and for measuring such resemblances in a rough-and-ready way, a coefficient of correlation may legitimately be used without too severe an insistence upon these particular postulates. After all, the use of a product-moment coefficient does not *necessarily* imply that the correlated variables are normally distributed.

(2) But I have no wish whatever to exclude a preliminary standardization in every case. Indeed, my *Memorandum* ([93], pp. 261 *et seq.*) dealt in some detail with the type of case in which standardization seemed to me to be essential. Where considerable manipulation of the matrices was required, I suggested a slightly modified procedure involving two requirements: (*a*) $\Sigma x = 0$, and (*b*) $\Sigma x^2 = 1$ (where x denotes the raw marks). I termed this 'unitary standard measure' to distinguish it from the more usual reduction to terms of the standard deviation, i.e. putting (*c*) $\Sigma x^2/N = 1$. Stephenson explicitly adopts this modification ([96], eq. (1) to [4]); but he appears to think that the application of (*b*), which converts covariances into correlations, will thereupon abolish all relations between the two sets of factors obtained by correlating traits and persons respectively. Actually, as we have seen in the preceding paper, the matrix equation, expressing those relations in their simplest form, is merely modified by the addition of a diagonal pre-factor.

In my view the chief obstacle that the generalization of my argument has to meet is due, not to (*b*), i.e. to the conversion of covariances into correlations, but to (*a*), i.e. to the conversion of raw scores into deviations about the mean, and of unadjusted product-moments into covariances. The great difficulty in psychology is not that the unit, but that the origin or zero point,

remains so dependent on an arbitrary choice. This difference of zero point, or (as it may be called) of average level, operates as a general factor, and has to be eliminated, either statistically by what is virtually partial correlation (crude or precise), or experimentally by selecting a sample in which the averages are approximately the same for every individual. In my article I chose the latter. My chief reason was that, by avoiding differences in the irrelevant 'general factor' from the outset, and working with a relatively homogeneous population, the graded differences of type would stand out more conspicuously.

Stephenson does not apparently criticize this choice: for he himself has proposed it as being "the simpler case to examine."[1] In his subsequent calculations, too, he commonly treats the general factor for persons, introduced by differences between the averages for the traits, as non-existent or at any rate as negligible. Each mental type, he believes, is so sharply differentiated from the rest that, with his modified technique, he can proceed at once to extract the relevant type-factors, without needing to eliminate a general

[1] [97], p. 348. Here, in his own account of Q-technique, he proposed first to reduce the scores to deviations about the means of the *columns* and then to reduce the new deviations to deviations about the means of the *rows*. He then assumed, for purposes of a simplified exposition, that with such a mode of calculation it was "possible, although unlikely," that a set of scores could be obtained which "will satisfy both conditions (3) and (4)" (viz. $\Sigma_i x_{ij} = 0$ and $\Sigma_j x_{ij} = 0$), i.e. that the marks for each person and for each test should simultaneously add up to zero. This was, in fact, the mode of 'standardization' I had described in suggesting an analysis by covariances, though it differs from the mode of standardization ultimately preferred by Stephenson (*Psychometrika*, *loc. cit. sup.*).

The supposed difficulty to which Stephenson here alludes has been raised by more than one research student (e.g. by Miss Knowles in her 'Studies of Temperamental Traits by Q-Technique'). Obviously it is "possible" to write down an arbitrary set of figures which will satisfy both conditions (3) and (4); but, they ask, is not such a result "unlikely," except for a "few very special sets of figures"? Suppose we take an actual set of measurements and 'standardize' the figures by columns, so as to satisfy equation (3); we then proceed to restandardize the figures by rows; to do this, we subtract a *different* figure from each figure down the column: surely, it is argued, except for rare and unlikely cases, this must upset the original standardization which made that column add up to zero. Stephenson himself makes a somewhat similar point. The reply, of course, is that the column of averages subtracted themselves necessarily add up to zero: consequently, although the individual figures in each column are changed, the total of the column remains unchanged. The result, therefore, is not "unlikely"; it is an inevitable result of the mode of calculation described.

factor first of all. Moreover, a doubly-centred table of residuals (i.e. a bipolar correlation table with the centroid of the whole system at the origin), very similar to my own, is regularly obtained by those who use a summation formula to calculate and remove the first (or 'general') factor; and, if we suppose that the elimination implicitly removes the first factor from the original measurements at the same time, we may regard the doubly-centred matrix of residual correlations as expressing correlations between the rows of a doubly-centred matrix of residual measurements. My preliminary selection of cases, therefore, merely carried out at the experimental stage what the factorist carries out at the statistical stage.

Nevertheless, several critics and correspondents[1] have questioned

[1] This point is made, for example, by Dr. P. R. Jameson, who advances another argument which I have not dealt with in the text. He writes: "Why adopt in work on temperament a method of selection and marking which would never be adopted with intellectual tests?" The 'method of marking' to which he objects is that which involves grading traits for persons instead of persons for separate traits; but in point of fact this *has* been used with 'intellectual tests': tests like those in the Binet scale, for example, have been ranked in order of difficulty for groups of children (or for individuals) to show that a particular individual or group—boys, for example, as distinct from girls—are of (say) a verbal rather than a non-verbal type, etc. As for the 'method of selection,' the answer is that, were we trying to establish the existence of types in some intellectual or cognitive sphere—for example, the distinction between a verbal and a non-verbal type—we should surely (so far as possible) avoid choosing for our experiments a population that was highly heterogeneous in regard to general intelligence: such special types would be far more easily detected when we look for them in groups which are otherwise fairly uniform and whose intelligence is upon much the same general level throughout, than if we had a highly mixed sample containing both feebleminded children and scholarship winners (cf. [35], p. 63). In all branches of science it is a popular methodological maxim that, when an effect varies as a function of several causes, our inquiries should be so designed as to keep all but one of the causes approximately constant: to verify Boyle's law, we keep the temperature constant; to verify Charles' law, we keep the pressure constant. No doubt, this ideal of varying one condition at a time is commoner in the simplified expositions of the elementary textbook than in the practice of advanced research. But in a first preliminary account, which was all my article aimed at, a simplified exposition seemed the first essential.

To the statistical critic may I point out that the principle of first deducting averages of columns and of rows is identical with that adopted in the 'analysis of variance' when calculating 'discrepance'? The object of my own calculations might therefore be described as carrying the analysis of variance and covariance to a further stage by seeking possible factors in the table of discrepance itself.

my procedure, or at any rate my wish to generalize its results. Hence in what follows I propose to fulfil my promise, and shall attempt an analysis of an unselected, non-homogeneous sample. To make a detailed examination of the simplest case seemed desirable first of all, because more than one writer had emphatically denied that, even in that case, any functional relation could be found, much less a relation so simple as mathematical identity.

(3) Stephenson further objects to the use of a 'multiple-factor' method of analysis based on the method of least squares. To me this appeared essential in analysing correlations between persons, at any rate in theoretical work, since no other formulæ would permit the algebraic proof of the identity of P- and T-factors. Stephenson, however, maintains that this procedure "leads to artificial factors rather than genuine factors," which he holds can only be reached "by a Spearman technique." Accordingly, as we have seen, he prefers to substitute a 'two-factor' procedure. The 'two-factor technique,' however, was designed solely for hierarchies in which there are no indications of any other factor besides one 'general' factor and n specifics. But in correlating persons the type-factors nearly always appear as non-specific group-factors, or as bipolar 'general' factors over and above the first: in fact, as Stephenson's own more recent work appears to show, and contrary to what is supposed to obtain in correlating tests, multiple factors are the rule and not the exception. For rough practical purposes I see no objection to substituting a summation formula for the least-squares formula as a method of multiple factor-analysis; only I regard it as no more than a quick and simple procedure for reaching first approximations to the true figures. And, as a matter of fact, in most of the earlier preliminary studies on correlating persons, both by my research students and by myself, the summation formula was employed. However, the method of simple summation, as applied by Stephenson, leaves larger residuals than are obtained with the method of least squares; and, as Davies has shown, some of the residuals which Stephenson treats as significant, and on which some of his type-factors are based, become almost negligible when recalculated by the method of least squares [130].

(4) Finally, instead of applying what I have termed the 'general-factor method' ('method b' of my *Memorandum*) and analysing the correlation table as a whole, Stephenson prefers, as a rule, to use what I have called the 'group-factor method' ('method a'). The former, he considers, must inevitably lead—like the correlation of traits—to mere 'bipolar types' which "at most can be regarded only as extreme cases—opposite tail-ends of a single normal and

continuously graded distribution of persons"[1]—and this he regards as tantamount to denying the existence of types; the latter method, on the other hand, assumes the presence of separate and discontinuous types, which is apparently the only type recognized by Q-technique (cf. [138], p. 277, § 11).

The way in which types (or the tendencies underlying them) are distributed is a question which in my view can only be settled by an actual survey; such a survey I propose to report in a later chapter. It would seem, however, that, if we are to regard the secondary factors as group-factors, that of itself really invalidates the two-factor formula on which Stephenson claims to base his modified technique. But if the strict algebraic formulæ deducible for the group-factor procedure are adopted instead (and, I gather, Stephenson is quite ready to accept them: see below, pp. 392 and 415), then it can be formally shown that the results of the group-factor method have a very simple relation to those of the general-factor method: as we have already seen, a bipolar factor can be regarded as expressing the difference between two positive group-factors; and what appears as a bipolar 'general' factor with one sample of tests (or persons) may appear as a pair of limited 'group'-factors with another.[2]

To meet these various points of criticism, I shall here remove the limiting conditions imposed in the previous research. First, instead of starting with a rigidly selected group, all on the same level for general emotionality, I shall take a more ordinary sample, which will consequently

[1] *Character and Personality*, IV, p. 295. In 'The Foundations of Psychometry' [96], p. 209, however, he speaks of the factors in 'Q-technique' as being 'each a *common* factor'; and in his more recent paper on *Q-technique* (p. 9, footnote 1) he says that, by the use of reference values "we are able to make an analysis substantially similar to those described by Holzinger as bifactor and as methods *a* and *b* by Burt," from which it appears that he would be prepared (as I should) to use either according to circumstances.

[2] See p. 309 above, and p. 390 below (footnote 1), also [116], p. 360. In an earlier discussion of the "two contrasted types distinguishable amongst the emotionally unstable" I myself treated and described the corresponding statistical components as "group-factors." Certainly, when we are dealing with a small and discontinuous sample of variables, composed chiefly of extreme cases, the group-factor method seems more appropriate from a mathematical standpoint. But can we from a psychological standpoint claim that such types are really discrete? It is no doubt tempting to postulate two positive factors, characteristic of each of the two groups, and to identify these two factors with a pair of specific causes—two different genes, two

be more heterogeneous in regard to the irrelevant and obscuring factors. Secondly, I shall base the factor-analyses exclusively on correlations instead of on covariances. Thirdly, instead of applying the somewhat novel and elaborate calculations of the 'least-squares' method, I shall be content with the shorter and more familiar technique—the 'summation' or 'centroid' method—used by Spearman, Thurstone, and Stephenson himself. Whether the group-factor method is also to be substituted will depend upon whether or not the correlations themselves (or their residuals) indicate the presence of such limited group-factors.

Even with all these modifications it will still be found that the same conclusions emerge as before. The removal of the more stringent conditions must naturally yield results that are somewhat less exact: with approximate methods nothing but an approximate confirmation can be expected. But the divergences will seldom be much larger than the errors due to sampling.

Yet once we have admitted these simplifications into the theoretical proofs, it would surely be foolish to return to more elaborate computations for the rough requirements of practical work. If, for example, the closer fit given by the method of least squares is discarded in analysing the covariances, why need we introduce that method when calculating 'true regression estimates'? We shall thus be

different glands, or two different psychoanalytic mechanisms, according to our interests and theories. If I had to hazard a double parallel myself, I should perhaps suggest looking for an adrenergic (sympathetic) and a cholinergic (parasympathetic) temperament respectively. But it might prove equally plausible to regard the 'inhibited type' alone as due to a positive factor (or rather set of factors) and to assume that the intensity of this factor varied continuously in degree, so that the 'uninhibited type' would be simply a highly emotional personality developing almost naturally in the *absence* of such inhibiting factors (see below, p. 434). However, all such fancies are far too speculative to offer a means of deciding between two modes of factorization. They may perhaps indicate feasible lines of research; as explanatory efforts they are bound to lead to over-simplification. I myself prefer to look upon the phenotypes as complex resultants of a disturbance in the relative balance or equilibrium between these numerous underlying causes; and accordingly, in the absence of fuller knowledge, I regard a bipolar factor as the safer mathematical conception.

led to inquire whether the arduous arithmetic imposed by the full P- and Q-techniques cannot be legitimately abridged. Suppose, for instance, that, having determined what the types are, we propose to examine their frequency-distributions: we shall require to 'measure for their approximation to the type' not 'fifty to a hundred persons as variables,' ([92], p. 18), but several hundreds; and to work out the intercorrelations between all these persons, and then factorize the figures, would require the assistance of a full-time computer and scarcely seem justifiable for the crude observational measurements at our command. The ideal procedure, I must still insist, would be to correlate and factorize, not the persons (who may run into thousands) but the traits (which can be reduced to about a dozen). That, however, is precisely the procedure to which my critics now take exception as committing us at the outset to a denial of discrete types. Accordingly, one of the minor objects of this paper will be to show how the essential figures can be reached (at any rate within the limits of approximation accepted by those who adopt the summation method) by a far quicker and more direct approach.

CHAPTER XVII

ANALYSIS OF AN ILLUSTRATIVE GROUP

To test the effect of these various simplifications let us take, to begin with, a small but representative group: in the end it will be seen that the argument that I shall put forward may be generalized to any number of persons. With the students entering a particular training college year after year, and attending the classes in psychology, it was my practice to apply intelligence tests and obtain temperamental assessments, according to a prearranged plan. Here I shall select the little batch for whom the temperamental assessments show the highest reliability. It consists of twelve women only. Since my object is to illustrate methods rather than to demonstrate results, it is desirable that the group shall be small, so that the tables can be easily grasped by the ordinary reader: but at the same time it must contain a sufficiently varied set of personalities. It so happens (for reasons that are fairly obvious)[1] that both these requirements are met by this choice.

[1] To obtain satisfactory reliability coefficients for all the traits assessed it is necessary that every individual in the group should be thoroughly well known to at least two judges and that the judges should be the same for all: for this reason the group is necessarily small. Moreover, judges are far more likely to agree over a group that includes a number of well-marked and widely different personalities than over a more homogeneous collection. Although the group is small and varied, it is fairly typical and well-balanced. Evidently, if we are going to make an analysis into bipolar factors and are not taking the whole universe of traits and persons, then both the traits and the persons selected must be, so far as possible, symmetrically balanced towards the two poles. There seems nothing exceptional or unreasonable in this requirement: it merely ensures that the design of our experiment shall be appropriate to the statistical procedure that we propose to employ—a requirement on which statisticians themselves are the first to insist. I may add that assessments for much larger groups have been correlated and factorized by several of our students in the course of their researches or statistical exercises: I hope that the most recent of these inquiries will shortly be printed with detailed figures. Unfortunately, it is not often

The initial assessments were based, not upon impressionistic estimates of character-qualities, but upon observation of actual behaviour, recorded in conformity with an abridged questionnaire; these were grouped together and combined into quantitative measurements for each of eleven primary emotional traits according to the scale of rating which I have commonly used for such work. The resulting measurements are given in Table I. For every trait the mean is zero and the standard deviation unity: [1]

practicable to publish huge tables of correlations even when they have been assiduously calculated.

[1] The scale and general method of assessment have been described more fully elsewhere (see *The Young Delinquent*, p. 417, *The Measurement of Mental Capacities*, pp. 29 f., and [53], pp. 61 f.). Ideally the mean and if possible the standard deviation of such assessments should be those of the total population. In practice, particularly if the group is large, and selected as a random sample of the total population, it may, as a rule, be assumed that, with an exact set of ratings, the mean and standard deviation of the group will approximately coincide with those of the total population, and that, so far as they do not, the empirical figures should be adjusted. If we do not make such adjustments, we may be introducing an irrelevant and probably spurious correlation of the type discussed below (cf. pp. 421 f.). Here, since we are required to work with correlations rather than covariances, the figures have been reduced to terms of the mean and standard deviation of the little group of twelve. Obviously, to make such adjustments on the basis of figures obtained from a small group only may seem somewhat precarious. Actually, however, the recalculation involves no very great change; and in any case our purpose here is a simple arithmetical demonstration only; we are using a population of 12 instead of one of 120 or 1200, merely because so huge a table could scarcely be printed in full or studied with ease. The reader, therefore, is asked to bear in mind throughout that the arithmetic is designed to apply to a far larger group.

It may perhaps seem out-of-date to adhere to a classification of emotional impulses, put forward by McDougall 30 years ago and much criticized since. I do not for one moment dispute that a better could nowadays be devised. Indeed, in the latest of the schedules cited in the foregoing paragraph, the reader will find that several further traits or rather headings have been added ('appetite,' 'possessiveness or acquisitiveness,' 'talkativeness,' etc.) which do not appear in the present list: the peculiarities of behaviour placed under those headings were not altogether ignored, but were for the most part distributed under other headings. An acceptable classification, however, should itself be based on a correlational study of the component behaviour-tendencies; and I have made many attempts to improve my own analysis by such means. Meanwhile, the reason for preserving the original scheme has been simply to keep the data gathered throughout the whole period comparable, until a complete revision can be attempted.

TABLE I

Assessments for Emotional Tendencies

Persons:	A	B	C	D	E	F	G	H	I	J	K	L
Traits.												
Sociability . .	1·90	1·56	— ·60	— ·24	·35	·24	— ·12	·48	— ·60	—1·07	—1·90	·00
Sex	1·54	1·35	—1·12	— ·04	1·50	— ·09	—1·69	·25	— ·62	— ·85	— ·46	·23
Assertiveness . .	1·95	1·00	—1·04	·09	1·11	— ·10	— ·97	·42	·20	— ·77	—1·76	— ·13
Joy	1·66	1·02	—1·32	— ·30	1·41	— ·06	—1·10	·45	·02	—1·05	—1·25	·53
Anger . . .	1·52	·78	—2·03	·39	·94	·23	—1·33	·64	·43	·08	—1·09	— ·56
Curiosity . . .	1·21	·23	—2·07	·05	1·22	— ·14	— ·50	— ·41	— ·76	— ·95	·54	1·58
Fear . . .	1·24	·77	—2·48	— ·09	·56	·04	— ·83	— ·21	— ·04	·26	— ·68	1·46
Sorrow . . .	1·26	— ·33	—2·14	·33	— ·98	·14	·98	— ·98	— ·46	1·20	— ·05	1·03
Tenderness . .	— ·32	— ·35	—1·57	— ·81	1·17	— ·05	— ·91	1·32	·15	—1·06	·66	1·77
Disgust . . .	·60	— ·04	—2·29	— ·77	·93	— ·17	— ·65	1·32	·00	— ·12	— ·41	1·60
Submissiveness . .	— ·79	·63	—1·70	—1·30	— ·16	·07	— ·95	1·54	·59	— ·22	·95	1·34
Average . .	1·070	·602	—1·669	— ·245	·732	·010	— ·734	·438	—·099	—·415	— ·495	·804

this is virtually the effect of the scale of rating, which, in accordance with the generally accepted principle, arranges that the figures shall be in standard measure. It will be seen at once that the means for the different persons are no longer identical: they range from $-1{\cdot}67$ to $+1{\cdot}07$. The first condition laid down above is thus fulfilled. We shall see later on how far such differences in average or general emotionality obscure the smaller differences of type which form our immediate interest here.

A. Correlations between Persons

Since we have agreed that the allotment of certain persons to one and the same type is to be decided by the resemblances which those particular persons show to each other, the first step will naturally be to estimate the amount of that resemblance by means of correlation. Accordingly, we begin by correlating the several persons. The coefficients are shown in Table II. The bipolar arrangement[1] which was noted in the mixed group of delinquents and neurotics reappears in this normal sample, though not perhaps quite so flagrantly. We see that the first six (*A* to *F*) all tend to resemble one another—*A* and *B* have positive correlations with every member of this sub-group: similarly the last six (*G* to *L*) resemble one another, though with them rather more of the correlations are low or negative: but in the main this second six (*G* to *L*) tend to show negative correlations with the first six (*A* to *F*). The negative correlations are as large as the positive. Hence we must look, not for two group-factors, but for one bipolar factor.

We have next to inquire how these resemblances and

[1] Although Stephenson dislikes my introduction of a bipolar *factor*, he himself admits the existence in his own correlation tables of a bipolar *pattern* (cf. e.g. [92], p. 302; [97], p. 349). Another critic suggests that in my previous group the bipolar division into antithetical sub-classes was due solely to the fact that I had actually started by combining two distinct sub-classes into one, namely, the extraverted delinquents and the introverted neurotics. He overlooks the fact that I had already obtained similar bipolar patterns with large and entirely random selections of the population—150 children and 170 adults (see [30], p. 695, Table II).

TABLE II

Correlations between Persons, and Saturation Coefficients for Persons

	A	B	C	D	E	F	G	H	I	J	K	L	Total.
A	[1·0000]	·5535	·3156	·8236	·2309	·1323	·0508	—·6030	—·4779	—·0733	—·8308	—·6825	5·6726
B	·5535	[1·0000]	·6553	·1801	·3083	·2414	—·4563	·0175	—·1013	—·4324	—·6729	—·7029	5·2869
C	·3156	·6553	[1·0000]	—·0424	·2497	·1126	—·1956	·1958	—·1700	—·6737	—·4880	—·5606	4·1829
D	·8236	·1801	—·0424	[1·0000]	·0255	·2496	·1895	—·7953	—·4239	·2686	—·4892	—·5517	4·0383
E	·2309	·3083	·2497	·0255	[1·0000]	—·5780	—·7811	·2736	—·0312	—·7511	—·2142	—·1587	2·8992
F	·1323	·2414	·1126	·2496	—·5780	[1·0000]	·3207	—·1325	·1149	·3750	—·2641	—·4583	1·2021
G	·0508	—·4563	—·1956	·1895	—·7811	·3207	[1·0000]	—·5155	—·3810	·5315	·0812	·2566	1·8449
H	—·6030	·0175	·1958	—·7953	·2736	—·1325	—·5155	[1·0000]	·6547	—·3931	·1511	·0872	2·0282
I	—·4779	—·1013	—·1700	—·4239	—·0312	·1149	—·3810	·6547	[1·0000]	·1367	·0785	—·0538	2·5246
J	—·0733	—·4324	—·6737	·2686	—·7511	·3750	·5315	—·3931	·1367	[1·0000]	·1642	·1202	2·8465
K	—·8308	—·6729	—·4880	—·4892	—·2142	—·2641	·0812	·1511	·0785	·1642	[1·0000]	·7678	5·2022
L	—·6825	—·7029	—·5606	—·5517	—·1587	—·4583	·2566	·0872	—·0538	·1202	·7678	[1·0000]	5·2926
Total[1]	5·6726	5·2869	4·1829	4·0383	2·8992	1·2021	1·8449	2·0282	2·5246	2·8465	5·2022	5·2926	43·0210
Saturation Coefficients	·86485	·80605	·63773	·61569	·44202	·18328	[—]·28128	[—]·30923	[—]·38490	[—]·43398	[—]·79313	[—]·80692	

[1] In calculating the absolute values for the totals, the signs of the top right and bottom left sub-matrices—from A to F and G to L—have been reversed. The original figures were calculated to eight decimal places, and have been rounded off for printing.

dissimilarities may be expressed in terms of one or more factors instead of a dozen or more traits. Our second step, therefore, will be to carry out a factor-analysis of the correlation table, so as to determine the saturation coefficients for the several persons, and the size of the subsequent residuals.

The method, we have decided, is to be that of simple summation. In applying it, we at once encounter two slight difficulties. First, what figures are we to insert in the leading diagonal in the place of the self-correlations or so-called 'reliability coefficients'? Different investigators, as we have seen, follow different plans, at any rate for calculating the mere saturation coefficients. Stephenson, for his modified Q-technique, before correlating persons, assumes that "the variates are first standardized . . . so that for each person the following holds: $\Sigma x^2_a = \ldots = \Sigma x^2_n = 1$," i.e. he adopts what I called 'unitary standard measure' ([96], p. 197, eq. [2]). This would permit us to regard the ensuing correlation table, if we like, as a table of variances and covariances *with the variances for persons always put equal to unity*. If only for its extreme simplicity, let us provisionally adopt this proposal here. Were we dealing with small tables, like those obtained by correlating traits, such differences in procedure might certainly introduce perceptible differences into the results: but, when we correlate persons, we shall generally be dealing with tables so large that any change confined to the leading diagonal would be almost entirely swamped in summation with the rest of the column: so that the problem is of minor importance.[1]

The second difficulty is caused by the numerous negative coefficients. In correlating traits these seldom appear in the table of observed correlations. Hence those who have simply carried over Spearman's methods and formulæ from correlating traits to correlating persons have found themselves in some perplexity. Once again, the usual course has been to omit them. Thus, in keeping with 'method *a*' (i.e. the 'group-factor method'),[2] they usually work first with one positive sub-matrix and obtain a first factor, and then with the other positive sub-matrix to obtain a second, which (as they rightly point out) is roughly "the obverse of the first," sometimes entirely ignoring the negative correlations,

[1] On the problem here raised, see pp. 332, 395, and 474.

[2] [93], p. 306. As implied above, it is strictly applicable only when the figures in the neglected sub-matrix of residuals are virtually zero, i.e. insignificant as compared with the size of the probable error.

although the line of division is said to depend upon them[1]: in practice, however, the line of division between the four quadrants is rarely clear. But if we imagine the signs of the initial measurements of the last six persons (G to L) to be reversed, thus reversing the signs of their correlations with the remaining persons, we can apply the summation formula to the whole matrix at a single step. To restore the signs we must afterwards prefix a negative to the saturation coefficients of G to L. The whole series will thus be accounted for in terms of a single bipolar factor. (See below, pp. 458, 485.)

With these two points provisionally settled, it only remains to add up the figures in each column of correlations, and divide each total by the square root of the grand total for the whole table. The totals, the grand total, and the resulting saturation coefficients are shown at the foot of Table II. On squaring the saturation coefficients we find that the factor so indicated accounts for about 35 per cent. of the entire variance. The next factor would account for barely 17 per cent., and the third for only 9 per cent. But with a group so small it is hardly profitable to consider these further factors.

B. Resulting Factors and Types

The factor-saturations for each person, thus deduced from the correlations of his measurements with those of all the rest, may be taken as measuring the degree to which that person approximates towards the type for which the factor is responsible. As I have argued in previous articles,[2] *these saturation coefficients are simply coefficients of correlations indirectly calculated*: when obtained by an adequate

[1] E.g. [92], p. 22; cf. *ibid.*, p. 302, Table I (in this table R and S have only one or two negative correlations instead of twelve; and hence I should have put them into the matrix for Type I, not Type II. Actually, hardly any of the negative correlations are significant: so that a single general factor, with positive saturations throughout, would fit the figures almost as well).

[2] E.g. [101], p. 89. My proof related only to saturation coefficients obtained by the method of least squares. But, as a matter of fact, this view of the saturation coefficient seems also to be accepted by nearly all who have preferred the simple-summation formula for correlations between persons; it was adopted, for example, by Beebe Center and later by Stephenson.

statistical method, they specify the person's correlation with the ideal or hypothetical person—the 'key personality'—who represents the pure type.

In passing, it may be of interest to inquire what the main factor thus indicated can represent. In the schedule of temperamental qualities, used in this and other researches, gradings are incidentally requested, not only for 'general instability,' but also for the 'tendency towards inhibition or repression.'[1] On comparing the saturation coefficients just calculated with direct assessments for the latter quality, I obtain a correlation of — ·83. Provisionally, therefore, we might identify this factor with 'extraversion' defined loosely as the opposite of inhibition or introversion. However, these impressionistic assessments are not themselves very reliable: it is unanimously agreed that K and L are of an inhibited or introverted type and that A and B are uninhibited extraverts, but over the remaining persons there is an appreciable disagreement. We must therefore postpone this question until we can interpret the factor in the light of the traits from which it has been deduced.

In any actual investigation it would, of course, be highly desirable to base the saturation coefficients on a much larger group; and the reader will now easily see how laborious the method would be were it essential to calculate all the possible intercorrelations before we could arrive at an approximate assessment for the temperamental type of any particular person in the group. As the advocates of the 'Q-technique' point out, "in applications to estimations for introversion-extraversion and the like . . . perhaps thousands of such correlations will have to be dealt with in any major study" ([97], p. 35). This seems to put the 'Q-technique' entirely out of the question for the humbler purpose of practical measurement at psychological clinics or in mental hospitals, or for surveys on any large scale to investigate the frequency-distribution of types.

Saturation Coefficients without Factor-analysis.—May I therefore at this point introduce a simplified procedure which will enable us to reach virtually the same results as those of the 'method of summation'? In proving the formula on which this latter method is based, I pointed out

[1] *Measurement of Mental Capacities, loc. cit. sup.*, p. 29.

a curious algebraic corollary,[1] from which it follows that (under certain not unreasonable conditions) identical

[1] *Marks of Examiners*, p. 287. Or, as I put it in another paper, "*g* is simply the average of the tests" ([102], p. 176. Cf. also pp. 330 f. above). This is my justification for determining 'true marks' by simply averaging ranks for the separate tests (e.g. in tests of æsthetic appreciation, etc.) and the order of 'general preference' by simply averaging ranks for individual preferences (e.g. in determining preferences for school subjects: cf. Burt, *ap.* Board of Education's *Report on the Primary School*, 1931, pp. 277–8; Pritchard, *Brit. J. Educ. Psych.*, V, 1935, pp. 15 *et seq.*): the minor factors included in such individual ranks or standard measurements are more or less cancelled by the process of averaging. Stephenson raised strong objections to this principle in his paper on "A New Application of Correlation to Averages" (*Brit. J. Educ. Psych.*, VI, pp. 43–57), urging that the opposite types produced by the minor factors are not cancelled, but ignored. Here, as elsewhere (so he insists), "a factor. . . should be clearly distinguished from a mere average" ([98], p. 357). It is, however, easy to show that the results obtained by formally eliminating the type-factors are virtually (if not precisely) the same as those that are obtained by the simpler and commoner procedure, and that these latter are usually as exact as the data will warrant (cf. Davies' reply to Stephenson [30]).

As was indicated in both the passages just cited, the 'corollary' mentioned in the text would seem to apply equally to Thurstone's 'centroid method,' since his cardinal equation is the same as my own (*loc. cit.*, eq. xxv): but when it comes to determining the factor-measurements, Thurstone himself, like Spearman, employs the method of least squares. The differential weights, however, derived by the method of least squares, although no doubt in theory supplying results of superior accuracy and validity, make little practical difference to the final averages or sums.

I may add that in the fuller version of my 1935 *Memorandum* I showed that a more general form of the corollary could be reached if the variances of the measurements to be summed are made equal, not to each other and to unity, but to the communalities of the tests or traits, and if the diagonal elements are made equal to the communalities instead of to unity. From a practical standpoint, however, the calculation of type-factor measurements on this basis would be more laborious and not less, and the gain in accuracy is all but negligible.

To the mathematical reader an apology is perhaps owing for the long attempt to prove in words a proposition which may seem obvious to him as soon as it is expressed in algebraic form. My excuse is, first, that several statistical investigators (including those who have used the summation formula most frequently) have questioned the proposition; and secondly, that a verbal argument and an arithmetical example will make the result plausible to numerous readers who either cannot follow the algebra or cannot accept the algebraic premisses in the precise form which is necessary for the simple mathematical proof.

figures may be attained without calculating the inter-correlations between the several persons or applying any factor-analysis at all.

The detailed proof (omitted in my *Memorandum* in its published form) has been given in the preceding paper. The following principle results. With the summation method and with unity in the diagonal cells, the factor-measurements are given by the unweighted sums of the measurements that have been correlated. Hence, for any given person, " the sum of his correlations is proportional to his correlation with the sums " ; consequently, to obtain his saturation coefficient for the first (or ' general ') factor, we have merely to correlate his marks with the sums or averages of all the persons' marks ([115], p. 164 : it is, of course, assumed that all marks or measurements are in standard measure).[1]

The same principle may be extended to other factors. Thus, to obtain a person's saturation coefficient for the second (or ' type ') factor we have merely to subtract the averages from the original marks of the several persons, standardize the residuals, and then correlate that person's residuals with the sums or averages of all the residuals, first reversing the signs of half the persons to make summation possible. In abstract terms the procedure may sound a little complicated ; but the arithmetic, it will be seen, is

[1] This is the principle underlying the rough-and-ready method that I have termed ' factor-analysis by simple averaging.' Suppose, for example, m_{ij} represents j's performance in the ith test, measured, let us say, in terms of speed. Then the persons' factor-measurements for the general factor common to all the *tests* (i.e. each person's general or average speed) may be taken as $\sum_i m_{i1}/n$, $\sum_i m_{i2}/n$, . . . etc. ; similarly, the factor-measurements for the general factor common to all the *persons* (i.e. the general or average difficulty of each test) may be taken as $\sum_j m_{1j}/N$, $\sum_j m_{2j}/N$, etc. Saturation coefficients for tests (or persons) can be calculated by correlating the observed measurements for each row (or column) with the averages for all the rows (or columns). For the observed table of measurements, a matrix of rank one can be fitted from the marginal totals by taking the estimated m_{21} as $= \dfrac{\sum m_{i1} \times \sum m_{2j}}{\sum\sum m_{ij}}$; and similarly for the other estimated m's. A table of residuals can be obtained by subtraction ; and (after reversing the signs where necessary) may be analysed for secondary ' bipolar ' factors in the same way.

exceedingly simple. Accordingly, partly to make the principle plain and partly because the identity of the results has been questioned, it will be worth while to exhibit it in a specific instance. Incidentally, the process will supply us with a far clearer notion of what the so-called saturation coefficient for persons really means.

What is implied in the procedure I am proposing may perhaps best be understood as follows. Let us recollect that, when we were following the old method of correlating traits with a view to discovering specific factors and types, the great difficulty that always confronted us from the outset was to eliminate the differences in average or general emotionality. The recognized statistical device was to apply partial correlation; this was equivalent to estimating each person's general emotionality by a *weighted* average of his measurements for the separate emotional traits, and deducting this estimate. Now, in statistical work it is a familiar experience that a weighted average seldom differs greatly or even significantly from an ordinary or unweighted average; and when we turn to correlate persons, we regularly begin by deducting the *unweighted* average for each person from his several measurements before we convert them to terms of the standard deviation.

Let us therefore make this step explicit. Ignoring any suggestions for weighting, let us accept the plain average of each person as a fair assessment of his general emotionality, and eliminate all individual variation in that respect by the simple process of subtraction. Table III gives the result of this elimination, the remainders or residuals being reduced to standard measure. At this point the computer might go on to cross-multiply all the columns, each with each, and so obtain correlations between persons as before. This, however, is not required. Instead he is asked to regard these standardized residuals as forming approximate measurements for each person in the second or more specific factor which remains after his general emotionality has been deducted and is responsible for the classification into temperamental types.[1] Accordingly, to obtain an indication of each ideal type

[1] Actually, it may be said, the measurements must include not only errors, but also measurements of any further factor, for we have only eliminated the first: to eliminate the third and later factors we should have to apply the method of least squares, which alone yields uncorrelated saturations. Here, however, we are content to suppose (as is assumed by the summation method of factor-analysis applied by Spearman, Thurstone, Stephenson, and others) that the influence of the further factors will virtually cancel out when the figures are averaged.

TABLE III

Assessments expressed as Multiples of the Standard Deviation for Each Person

Persons:	A	B	C	D	E	F	G	H	I	J	K	L	Total.[1]	Total (in s.m.)
Traits.														
Sociability .	·977	1·551	1·888	·009	— ·535	1·672	·917	·057	—1·168	— ·945	—1·533	—1·029	9·263	1·412
Sex . . .	·553	1·211	·969	·412	1·076	— ·727	—1·430	— ·255	—1·215	— ·628	·039	— ·735	7·718	1·177
Assertiveness .	1·036	·645	1·111	·674	·530	— ·800	— ·353	— ·025	·697	— ·512	—1·380	—1·195	5·964	·909
Joy . . .	·695	·677	·616	— ·112	·950	— ·509	— ·548	·016	·278	— ·930	— ·824	— ·351	4·676	·713
Anger . .	·530	·289	— ·637	1·278	·292	1·599	— ·892	·273	1·234	·713	— ·649	—1·746	4·418	·674
Curiosity . .	·165	— ·602	— ·708	·593	·684	—1·090	·349	—1·149	—1·541	— ·772	1·130	·992	·033	·005
Fear . .	·200	·272	—1·432	·311	— ·241	·218	— ·144	— ·878	·138	·972	— ·201	·839	— 1·398	— ·213
Sorrow . .	·224	—1·509	— ·831	1·157	—2·398	·945	2·562	—1·921	— ·842	2·327	·486	·288	— 5·312	— ·810
Tenderness .	—1·637	—1·541	·175	—1·139	·614	— ·436	— ·264	1·194	·581	— ·930	1·261	1·235	— 7·041	—1·074
Disgust . .	— ·553	—1·039	—1·096	—1·059	·278	—1·308	·125	1·194	·231	·424	·093	1·018	— 7·862	—1·199
Submissiveness .	—2·190	·046	— ·055	—2·126	—1·249	·436	— ·323	1·492	1·607	·280	1·578	·685	—10·457	—1·594
Correlation with Total . .	·86485	·80605	·63773	·61568	·44204	·18327	—·28128	—·30922	—·38490	—·43398	—·79313	—·80692		

[1] In calculating these totals the signs in the last six columns (G to L) have been reversed.

itself, all that is necessary is to average a sufficiently large series of such residuals for an appropriate selection of persons. To measure the ideal extravert, we could average the figures for the half who appear to verge in that direction (here A to F); to measure the ideal introvert, we could average figures for the other half, who appear to verge in the opposite direction (here G to L). This, in effect, is the procedure followed by those who prefer to deal with the two groups separately. If, however, we assume that introversion is the exact antithesis of extraversion, we may reach a much better measurement for both by taking *all* the figures for the group, first reflecting or reversing the signs for those belonging to the second sub-group. The column of averages or sums thus calculated will represent the theoretical measurements for the ideal extravert type. The simple totals are appended in the last column but one of Table III, and are reduced to ordinary standard measure in the last column of all. To obtain analogous measurements for the ideal introvert we have merely to reverse the signs.

At this point the theorist may be tempted to urge that we are basing our figures for the type largely on persons who do not really represent it. If the types were sharply separated, he will say, the procedure might pass: but if a large number in the group are more or less intermediate between the two, ought we not to omit them altogether or at least give them a very low weight? Before we reply to this objection, however, let us accept the results provisionally, and see how closely (or how remotely) the different persons actually correlate with them.

Taking each person in turn, then, we correlate his residual marks with the sums or averages of all the residual marks, calculated as shown in Table III. The correlation thus obtained should express the degree to which each person approximates to the extraverted type. How far do the figures obtained by this simple method agree with those obtained by the previous and far more lengthy procedure? The correlations are given at the foot of Table III. It will be seen that each correlation is *not merely an approximation to, it is identical with, the saturation coefficient for the same person as reached by the more elaborate method of intercorrelating all the persons and applying a factor-analysis by the summation method.*[2]

[2] To exhibit the exact arithmetical identity I have carried the calculations to far more decimal places than the data would otherwise warrant. The

I have demonstrated this result for a small group of twelve persons only. The method can obviously be generalized to a sample of any size; and, as I have already insisted, a safe measurement of the ideal type could only be attained when the figures were derived from a fairly large sample. But, whatever the size of the group, the mathematical reader will perceive that it is, in fact, an algebraic necessity that the final figures obtained by the two alternative methods shall be exactly the same. If I am right, it follows that the lengthy computations proposed by Q-technique —the preliminary calculation of many hundreds of correlations between persons, the elaborate factor-analysis of a table of coefficients, and the final calculation of saturations or regression coefficients—are really unnecessary for the measurement of resemblances to types: the same figures, or virtually the same, can be obtained by the simple procedure outlined above.

Determination of Types.—We may now inquire what is the apparent psychological nature of the types and factors to which these modified methods have led us. Are they essentially the same as were found in the set of more abnormal cases referred for clinical examination and treatment? Are they the same as we should discover in a larger sample of the general population selected wholly at random? If not, are the fresh factors and the new types due to the altered mode of selection or to changes in statistical procedure?

The obvious way to determine the nature of the two antithetical types would be to study the marks or weightings allotted

slight discrepancies in the fifth decimal place are, of course, due to the rounding off of the previous figures. The student who desires to verify the identity has merely to multiply the figures in any one column with those in the last column (totals reduced to standard measure) and divide the sum of the products by eleven.

In passing, I should like to stress the practical importance of the principle that has been twice invoked in the foregoing argument. It maintains that, unless the weightings are widely different, and unless the weights themselves and the measurements to be weighted are highly exact, we may obtain very reasonable approximations to the hypothetical measurements for *any* general factor by taking simply the unweighted averages of the appropriate empirical measurements just as they stand (cf. [93], p. 310, para. 7).

by each factor to the eleven emotional traits. However, this proposal raises several questions about which different investigators are not altogether agreed.

To discover the nature of the 'positive type' (as we may call it for the moment) the simplest method of all would be to compare the figures for each trait, as shown by the final column of sums in Table III; to discover the nature of the 'negative type' we should take the same figures with the signs reversed. But, since the calculation of these figures rests on an argument which may not be generally accepted, it will be better first to inquire what other means of determination may be available.

For those who treat the problem of comparing and correlating persons as merely another case for the application of Spearman's technique (first 'inverting' or, as I should prefer to say, transposing the initial matrix of measurements) the most logical procedure would be to employ Spearman's regression equation or 'weighting formula'—a method which, as it happens, is based on the principle of least squares: only, instead of constructing a regression equation to get factor-measurements for the persons, we shall now, of course, construct one to get factor-measurements for the traits.[1] The calculations would be long; but they were apparently [2] carried out by Beebe Center; and his methods appear in the main to have been followed by Stephenson in his recent article. At all events Stephenson himself states that the figures he deduces to describe his two temperamental types should be regarded as 'true regression estimates' ([92], p. 304). Elsewhere, however, he explains that "the person with the highest saturation is the most useful for predicting (by correlation) the [measurements of the] type to which all persons of one factor are approximating"; and to save labour it would seem that he generally employs simple inspection in order to pick out the most typical person or persons in his batch.[3] If so, to single out one or two persons from a small batch of twenty or thirty would appear a somewhat precarious method of setting up a standard.

[1] [56], eq. 31, p. xx. An illustration of the calculations as adapted to obtain factor-measurements for traits from correlations between persons was given in my previous paper ([114], p. 175).

[2] [80], p. 207. (Actually Center's final values are expressed in terms of simple ranks.)

[3] [97], p. 351. In the only case in which figures are actually given for predicting' or describing the type ([92], p. 303) the figures are simple integers from 0 to 10, and obey his prescribed frequency distribution: so that they are presumably merely the assessments of the one most typical person in his batch of 21.

But, however applied in actual practice, the underlying principle, it will be noted, is to estimate resemblances between individuals and the 'type' first of all, and deduce the 'type' afterwards. With the method described in the previous paragraphs the order of procedure is reversed: we first determine the 'type,' and deduce the degree of resemblance to it afterwards. Moreover, with this method we use measurements from *all* the persons to determine the coefficients for the type, and give to each person an equal weight. At first sight this seems far better than using one person, or even one small group of persons picked out by simple inspection. Yet the method itself demonstrates in the end how widely the various persons differ in their resemblance to the hypothetical types. Does it not therefore follow that the principle of equal weighting was after all very inadequate? Would it not be better to weight the several persons differently according to their saturation coefficients as now obtained? We may readily agree. But, if we go on to revise the whole of our computations on this basis, we shall find that the correlations with the type as thus recalculated will not be the same as those with which we started out. We must therefore begin again with these revised saturation coefficients. Thus we shall be continually led to fresh coefficients and fresh weights: in short, to a spiral process of successive approximation. If we keep on we shall ultimately discover that, having started with saturations determined by simple summation, we are gradually approximating towards the saturations determined by least squares. That being so, would it not be much quicker to seek from the very outset a direct determination of the type by correlating traits instead of persons? For, with the method of least squares, the saturation coefficients for *persons* are really the correlations of those persons' measurements with the saturation coefficients for *traits* ([114], p. 178); and to reach the latter we must begin by correlating traits.

C. Correlations between Traits

But this brings us back to our central problem. Can we still assume that the factors obtained by correlating traits will be identifiable with those already obtained by correlating persons, even under the more general conditions that we have here laid down? For the ordinary reader, the most convincing way to answer this question will be to examine calculations for a small table of figures such as that given above, and then note how the principles involved are perfectly general.

TABLE IV

Correlations between Traits

	Joy.	Fear.	Sex.	Disgust.	Assertiveness.	Anger.	Curiosity.	Sociability.	Tenderness.	Submissiveness.	Sorrow.
Joy	[·853]	·758	·925	·694	·948	·828	·684	·828	·493	·270	·075
Fear	·758	[·812]	·697	·818	·658	·743	·809	·526	·524	·454	·609
Sex	·925	·697	[·624]	·580	·854	·791	·706	·725	·447	·252	·002
Disgust	·694	·818	·580	[·731]	·523	·610	·732	·378	·847	·722	·339
Assertiveness	·948	·658	·854	·523	[·967]	·874	·488	·883	·232	·047	·071
Anger	·828	·743	·791	·610	·874	[·922]	·490	·651	·294	·233	·224
Curiosity	·684	·809	·706	·732	·488	·490	[·865]	·383	·678	·389	·450
Sociability	·828	·526	·725	·378	·883	·651	·383	[·946]	·076	—·025	·106
Tenderness	·493	·524	·447	·847	·232	·294	·678	·076	[·783]	·815	—·007
Submissiveness	·270	·454	·252	·722	·047	·233	·389	—·025	·815	[·879]	·044
Sorrow	·075	·609	·002	·339	·071	·224	·450	·106	—·007	·044	[·850]
Saturation Coefficients:											
Factor i	·951	·898	·878	·830	·827	·824	·780	·671	·564	·412	·259
Factor ii	·233	—·159	·213	—·490	·483	·421	—·268	·508	—·667	—·685	—·104

Let us therefore proceed to compute the correlations between the various traits as measured in Table I, and then factorize this second correlation table. The correlations and reliability coefficients, calculated in the ordinary way, are shown in Table IV. Practically every correlation is positive; and it is at once obvious that the figures are tending towards as good a hierarchical order as we could expect with a group so small and traits so multifarious. Such an order implies, not only the presence, but the predominance, of a general factor common to all the traits: this first factor is, of course, virtually eliminated when we correlate persons. On eliminating its effects from Table IV, at least one other factor—possibly two—are discernible. The saturation coefficients have been calculated for all three,[1] and are given, rearranged for purposes of comparison, in the first three columns of Table V.

Here we are interested neither in the first factor ('*the* general factor,' as it is commonly termed), nor yet in the third factor whose nature and even existence here seem highly dubious. Our immediate interest lies in the second. This, if any, will be the factor that corresponds with that discussed above while correlating persons. The trait-weights for the two, as obtained by correlating traits and persons (with the 'abridged method' for the latter), must accordingly be compared. The figures are given in the second and fourth columns of Table V, headed t. ii and p. ii. The figures in the latter column are simply the

[1] The calculation was carried out in accordance with Dr. Stephenson's own description of his method. The procedure is based on Spearman's well-known summation formula. It differs from my own in implying a slightly different figure for the self-correlation of each test. If this implied figure is inserted, then a check with table-by-column multiplication shows that the saturations obtained do not differ greatly from those that would be obtained by the least-squares method *with these self-correlations*. The choice of figures for the self-correlations is largely arbitrary; and I myself should prefer to insert values that would give (i) zero correlations between the saturation coefficients for different factors and (ii) a total of zero for the bipolar saturation coefficients of the same factor. This would incidentally yield saturations for t. ii that were much closer to the figures given under p. ii.

TABLE V

SATURATION COEFFICIENTS AND FACTOR-MEASUREMENTS FOR TRAITS

	Correlating Traits.			Correlating Persons.			
				Abridged Method.	Average.		
	t. i	t. ii	t. iii	p. ii	p. i	p. ii	p. iii
Sociability	·671	·508	·093	·587	·69 *1*	·55 *1*	·15 *2*
Sex	·878	·213	·177	·489	·58 *5*	·19 *5*	·11 *4*
Assertiveness	·827	·483	·104	·378	·64 *2*	·51 *2*	·13 *3*
Joy	·951	·233	·225	·297	·53 *6*	·10 *6*	·34 *1*
Anger	·824	·241	—·109	·280	·59 *4*	·44 *3*	—·35 *10*
Curiosity	·780	—·268	·198	·001	·23 *11*	·25 *4*	·02 *7*
Fear	·898	—·159	—·317	—·089	·62 *3*	—·39 *10*	—·16 *9*
Sorrow	·259	—·104	—·798	—·337	·44 *9*	—·23 *8*	—·38 *11*
Tenderness	·564	—·667	·353	—·447	·51 *7*	—·12 *7*	·09 *6*
Disgust	·830	—·490	—·104	—·489	·30 *10*	—·25 *9*	—·08 *8*
Submissiveness	·412	—·685	·237	—·525	·48 *8*	—·52 *11*	·10 *5*

totals given at the end of Table III reduced to terms of the same standard deviation as those entered under t. ii.

It is true the two sets of figures are no longer absolutely identical. But absolute identity was hardly to be expected after we had decided to substitute a rough summation method for both calculations. Nevertheless, with the exception of one rather indeterminate trait—curiosity[1]

[1] In most of the other groups that I have investigated, 'curiosity' has appeared as definitely characteristic of the 'objective,' 'aggressive,' or 'extraverted' type. With this particular batch—students at the end of their adolescent period, it will be remembered—the conduct which gains a high mark for 'curiosity' consists of inquiring interests of an abstract and theoretical nature rather than of curiosity about other persons or about concrete things and practical situations: such curiosity is, intelligibly enough, associated rather with tendencies towards introversion.

My friend, Mrs. Milner, in her very remarkable book, has criticized my classification on the ground that 'disgust has usually an aggressive component, while sex is aggressive only in the male: in the female it is submissive' (*The Human Problem in Schools*, 1938, p. 106). My primary aim, however, was not to classify emotions, but to classify persons. I can only say that, so far as my data go, both males and females who are rated by observers high for sexual behaviour tend, on the whole, to be slightly more aggressive in their general conduct (though the correlation is quite low, especially in females). On the other hand, 'disgust,' I agree, as defined in the dictionary is often synonymous with indignation, and to that extent would be characteristic rather of

—the signs remain the same, and the magnitude of the figures is but slightly altered. The same seems to hold good of the third factor. We may, therefore, fairly conclude that, even when the group contains persons differing appreciably in their average or total amount of general emotionality, the remaining factors obtained by correlating traits can be regarded as identical for all practical purposes with the factors obtained by correlating persons. The minor differences are evidently the result of the differential weighting that has been implicitly introduced. This effect is diminished rather than increased when we apply the same method to larger samples. Hence we may, I think, legitimately assume that our conclusion can be generalized.

D. Resulting Factors and Types

It is interesting to note that in their concrete nature the several factors obtained from the small group of adults tally with those obtained from the larger groups of children [114].

The first factor here accounts for 56 per cent. of the total variance. Its saturation coefficients are positive throughout. The factor-measurements for the twelve persons differ but little from the averages given at the foot of Table I; and both correlate highly with direct judgments for 'emotional instability.'[1] We may, therefore, unhesitatingly identify this factor with 'general emotionality.'

The second factor accounts for 18 per cent. of the variance. Its

the aggressive type. But that was not the sense in which I used the word: following Shand and McDougall, I should class that type of behaviour under 'anger.' Disgust, as McDougall defines it, is more often inhibitive and repressive than assertive or aggressive.

[1] In discussing the results of factor-measurement as applied to patients in mental hospitals, Dr. Murdo Mackenzie has asked that "the alphabet of factors should be extended to include the factor of instability" ([91], p. 113). His request is already met by this factor of 'general emotionality.' We may perhaps loosely regard it as a central fund of emotional energy comparable to what certain psycho-analytic schools would called the libido. Elsewhere I have pointed out that the child in whom general emotionality is high becomes (unless his intelligence is correspondingly increased) essentially the 'unstable child'; and I have no doubt the same is true of adults. A more detailed account of the hypothesis of general emotionality will be found in *Character and Personality*, VII, pp. 238–254.

saturation coefficients are partly positive, partly negative. It is thus a bipolar factor. The saturations are positive for sociability, assertiveness, anger, joy, and sex; negative for submissiveness, tenderness, disgust, fear, sorrow, and curiosity. Both the twofold allocation and the relative size of the figures agree pretty closely with what was obtained by the simpler method here described (last column of Table III). We may infer that positive values of the factor indicate the aggressive, unrepressed, or extraverted type, and negative values, the inhibited, repressed, or introverted type. Incidentally, for purposes of practical diagnosis it is instructive to note that, of the more special characteristics assessed in the original schedule, talkativeness still shows a higher correlation with the second factor than any other.[1]

To the figures for the third factor little importance can be attached. In this group it contributes barely 10 per cent. to the total variance. The saturation coefficients are positive for tenderness, submissiveness, joy, sex, curiosity, assertiveness, and sociability; negative for sorrow, fear, anger, and disgust. The first list includes all the emotions that are attended with pleasure; the latter comprises the only four that are definitely unpleasant, the coefficient for sorrow being the largest of all. Put in this way the results are consistent with the suggestion that the third factor makes for cheerful or optimistic moods when positive, and melancholy or pessimistic moods when negative: if we prefer a

[1] With the homogeneous group it was ·78; with the present group, slightly lower, ·71. Statistically the difference is hardly significant, but it is in keeping with the complex nature of this symptom, as I have implied elsewhere. The importance of an assessment for talkativeness lies in the fact that the qualities most reliably assessed at an interview are those which the interview itself elicits (cf. [53], p. 64); but it may indicate one or more of at least three underlying tendencies. Speech and voice are no doubt influenced quite as much in their *quality* by specific emotions as facial expression is; but they are also affected in their *quantity* (or fluency) by (1) *general* emotionality, (2) special *inhibitive* tendencies, and (3) the *verbal* factor, usually regarded as a specific intellectual capacity, but as such probably of relatively small importance. Thus, the tongue-tied, taciturn child may be either unemotional, or inhibited, or (less likely) of poor ability in the use of words; the loquacious, garrulous child is usually both emotional and uninhibited (i.e. of the so-called extraverted or cyclothymic type). Now in the homogeneous group, differences in general emotionality were eliminated by the mode of selection; consequently, talkativeness, when present, appeared as a well-marked symptom of extraversion: hence the higher correlation. (Cf. *The Backward Child*, pp. 401, 547. The bearing of this on the use of performance instead of verbal tests with such children has been repeatedly emphasized, but is still too often overlooked.)

more technical name, we may call it a factor for *euphoria*. Actually however, the only coefficient of any considerable size is that for 'sorrow'; and I believe the true inference to be drawn is that two persons labouring under recent grief happened to find a place in this group at the time the estimates were made.

Since the group here described was small and deliberately chosen for illustrative purposes, no great weight could be attached to these points of agreement if they stood alone. Miss Knowles and others, however, have correlated and factorized assessments for much larger groups of our students, and have apparently reached very similar conclusions. Their larger tables will be found in their theses.

It will be as well to repeat that the factors thus demonstrated are, in the first instance, statistical components only: they are not psychological, physiological, or biological causes as such. It is perhaps possible that (as the more enthusiastic advocates of the method have claimed) the 'types discovered by factor-analysis' may be 'at bottom constitutional types, part of the individual's glandular inheritance'; but if so, factor-analysis alone could not demonstrate it. It has even been suggested that 'the extraverted temperament' may be 'due possibly to one gene and the introverted to its absence': I agree that the underlying conditions may, in part, be inherited as a 'Mendelian factor,' and have, indeed, given reasons for the belief;[1] but the characteristics of individuals as we observe them can be related to 'Mendelian factors' only in a very indirect and complex way: our statistical 'factors' are factors of a very different kind. Whether or not the groupings suggested by our crude data are at bottom partly the outcome of simple biological or biochemical causes, the statistical factors that express those groupings are, as we have seen, descriptive and predictive factors rather than causal factors: they seek merely to offer a quantitative picture of an individual personality, and to forecast his probable behaviour under various emotional stimuli. A different set of factors could easily be extracted which would describe the data almost as well: but for purposes of prediction they would be less simple, less convenient, less economical, and less accurate. What James wrote nearly forty years ago still remains true: "If we seek to place the emotions thus enumerated into groups, according to their affinities, all sorts of groupings would be possible—equally real and true. The reader may class the emotions as he will—as sad or joyous, sthenic or asthenic, natural or acquired, egoistic or non-egoistic, organismally or environmentally initiated, and what more besides. The only question would be—does this grouping or

[1] *Eugenics Review*, IV (1912), ii, pp. 189 *et seq.* Cf. pp. 8, 377, above.

that suit our purpose best?"[1] In the absence of a full insight into causes, the best classification for purposes of diagnosis and prognosis—though not perhaps of treatment—is a classification with a minimum number of divisions and a maximum number of deducible corollaries, each predictable with maximum probability. This, I maintain, is given by the statistical procedure here pursued; and the same procedure will alone enable us to give definite answers to still more pressing problems—how closely, for example, does the behaviour of this or that person remain true to the predicted type, in different situations, and in different moods, and at different stages of his life?

[1] *Principles of Psychology*, Vol. II (1901), p. 485.

CHAPTER XVIII

A REPLY TO CRITICISMS OF THE RESULTS

To what extent do the conclusions so far reached agree with those advanced by other investigators? Of non-statistical writers, Binet, Jung, McDougall, Kretschmer—not to mention numerous less-known names—have attempted detailed descriptions of the emotional characteristics of the two main types; and it would not be difficult to show that, so far as can be judged, these concrete verbal pictures correspond pretty closely with the abstract quantitative scheme that has been formulated above.[1]

[1] Jung himself does not directly define introversion or extraversion in terms of specific emotions. According to his original definition, "introversion is a turning inward of the libido [general emotional energy] whereby a negative relation of subject to object is expressed: interest does not move towards the object, but recedes towards the subject" (*Psychological Types*, 1924, p. 567). The implied distinction, as Jung has noted, is virtually the same as that previously drawn by Binet in contrasting his 'objective' with his 'subjective' types; and Binet certainly emphasized the emotional differences between his types. In my earliest researches on temperamental characteristics, before Jung had coined the more specific designations, I adopted Binet's vaguer adjectives to indicate the two opposite emotional dispositions, although I conceived the essential differences in terms of McDougall's 'primary emotions' (*Eugenics Review*, 1912, IV, p. 189). Strictly, the words 'objective' and 'extraverted' would seem best adapted to express the final attitude habitually adopted by the patient as a result both of inborn disposition and of post-natal experience rather than any quality of innate temperament alone. Of the two contributory influences inborn disposition is perhaps the more important: but it would be out of place here to discuss the wide but somewhat inconclusive evidence for such a view.

Of other writers I may merely note that the majority—e.g. Groos, Kretschmer, McDougall, James—like Jung himself, even when they do not *define* their types in terms of the primitive emotions, nevertheless largely *describe* them in those terms. Indeed, McDougall explicitly attributes the temperamental differences between the extravert and the introvert to 'differences in respect of the relative strength of the instinctive tendencies,' which in his opinion are identifiable with the fundamental emotions. Thus "E (the typical extravert) is extremely sociable: I (the typical introvert)

But a fuller and more precise correspondence can be established, if we turn to the results of those investigators who have used a statistical procedure. Here the work of Spearman and his school, together with Spearman's critical résumé of earlier researches, and the conclusions of Stephenson and others who have adopted Spearman's methods, afford the best basis for comparison.

I have endeavoured to meet Stephenson's criticisms of my methods. May I now briefly answer his criticisms of my results? As before, I believe that a little examination will show that the difference between us is far less than he imagines. The earlier conclusions of Stephenson and those who collaborated with him were based chiefly on tests applied to adults; but, as they point out, the factors discovered very closely resembled those that had been obtained in my own work with normal and abnormal children. The final summary of their results is given in the issue of *Character and Personality*, which also contains Stephenson's first experiments on correlating persons. For testing personality, however, and for studying clinical and temperamental types, they relied at this stage exclusively on R-technique—i.e. on correlating tests or traits. Since they were themselves the observers and the testers of every person in their group, this procedure was always open to them. I, on the other hand, had usually to collect the data for my clinical cases from *different* observers; and so had often been forced to start by correlating persons. Nevertheless, wherever it was practicable, I should still consider that 'R-technique' was the safer procedure, though Dr. Stephenson now seems to repudiate it.

shrinks from self-display; E's laughter is frequent and free; I is timid and shy . . . I is curious about scientific and metaphysical problems, gloomy, sensitive and even sardonic: his emotional expressions are repressed and restrained," etc. (*Outline of Abnormal Psychology*, pp. 436 *et seq.*, abridged). Holt epitomizes what is common to the two types of response by summarizing the one as 'aggressive or adient' and the other as 'avoidant' (*Animal Drive and the Learning Process*, 1931); and Katz regards 'these two fundamental reactions as the clue to extraversion-introversion' (in his chapter on 'Personality,' *ap.* Boring, Langfeld & Weld, *Psychology*, p. 510). For a general review of the whole subject, see Guilford and Brady, *Psychol. Bull.*, XXVII, 1930, pp. 96–107.

Their main conclusion, apart from minor modifications, is essentially the same as that put forward by Spearman in his well-known chapter on 'Types' ([56], pp. 41–54), namely, that nearly all the familiar antitheses observed by writers on temperament—extravert and introvert, objective and subjective, manic and schizophrenic, explosive and obstructed—are variants of the same theme, a factor (or possibly a set of factors) to be discovered by correlating tests, and identifiable with 'perseveration' or 'inertia,' as the "most fundamental of all the concepts involved." They suggest, however, several minor modifications. First, whereas Spearman had tentatively identified the essential type-factor with Müller's 'perseverative tendency' and the 'secondary function' of Groos, Stephenson holds it to be more definitely 'volitional,' akin to Webb's *w* ('persistence') and the 'determining tendency' of Ach. It is this volitional stability, in his view, that the alleged tests of perseveration ('*p*-tests') really measure, not the factor of 'inertia.' Secondly, in their later work they conclude (rightly in my view) that at least two factors have to be taken into account in diagnosing the temperamental characteristics of any individual patient, namely, *p* (or *w*) and *f*.[1] The possession of 'high *w*' and 'low *w*' leads to two distinguishable types, analogous to my 'stable' and 'unstable' types respectively.[2] The 'fluency-factor' (*f*) they identify with my second bipolar factor (*i*), namely, the tendency to sthenic (aggressive or assertive) emotions and asthenic (repressive or inhibitive) emotions respectively.[3] The possession of 'high

[1] The detailed evidence for these conclusions is set out in a series of 'Studies in Experimental Psychiatry' by the same authors published in the *J. Mental Science* of the same year.

[2] These I attributed to differences in a factor of 'general emotionality' (*e*). This factor, as I have elsewhere pointed out, would seem to represent the innate basis of Webb's *w* ([129], p. 254).

[3] The term 'fluency' is due to Hargreaves [58], who suggested that it might 'indicate inhibition or the reverse.' It would seem, however, to be an old factor re-named. Thus, Garnett identifies it with a factor that he had previously demonstrated and named *c* ('cleverness' [37]; cf. *Knowledge and Character*, p. 137 f.). I regarded it as nearly identical with 'quickness,' which, as measured by ordinary tests, I held to be a 'mixed or joint factor'—a resultant implying both high emotionality and

f ' or ' low *f* ' thus leads to a cross-classification into ' explosive ' and ' obstructed ' types respectively. These correspond with my ' unrepressed ' (or ' uninhibited ') and ' repressed ' (or ' inhibited ') types. They may be broadly identified, it is said, when due allowance has been made for differences in *p* and *g*, with the so-called ' extravert ' (' manic-depressive ' or ' cyclothymic ') and ' introvert ' (' melancholic ' or ' schizophrenic ') types of other psychologists. Hence, as Stephenson and his collaborators themselves pointed out at the time, their results and mine seem so far to be in close agreement.[1]

Now, as we have already seen, Stephenson's recent account of Q-technique, and even his further article in the same number of the same journal, are very difficult to reconcile with these earlier conclusions. When the latter were written, he was still of opinion that to correlate persons

lack of inhibition or repression. It is interesting to note that Hargreaves discovered the evidence for his *f*-factor in tests of imagination : this is quite in keeping with the close relation, found in our work on correlating persons, between differences in the factors of emotionality and inhibition and differences in types of artistic appreciation and production ([129], p. 294 f.).

[1] The minor differences seem attributable to the fact that *p*, *f*, and *w* are really ' mixed ' or ' non-fractional factors.' As indicated above, the results of such tests seem to depend upon a mixture or balance of at least two ' purer ' factors. To get back to these latter we must adopt a weighted combination (here a weighted subtraction) of the two. Thus the lines of classification given in these papers are drawn at slightly different points from my own. *w*, being in my view largely an acquired characteristic, must depend in part upon the second factor as well as upon the first. Hence Miss Studman rightly regards ' high *w* ' as largely characterized by ' inhibition.' Similarly she attributes ' exaggerated impulsion ' to ' high *f*,' whereas I should partly attribute it to ' high *e* ' (general emotionality).

Another of my former students, R. B. Cattell, in a slightly earlier research (thesis on ' Temperament Tests and Perseveration ' : cf. *Brit. J. Psych.*, XXIII, 1933, iii, pp. 308 *et seq.*), reached somewhat similar conclusions. He discusses my own terms (' Repressed ' and ' Unrepressed,' ' Sthenic ' and ' Asthenic '), and considers—I think with justice—that they may be taken to imply more than we are at present entitled to infer ; and consequently proposes to speak of ' Surgent ' and ' Desurgent ' types instead. His further *a*-factor—' optimistic and objective '—seems to correspond with the third factor found in my own research : he, too, is inclined to accept the view that this may be responsible for, or descriptive of, the minor contrast between the manic and the depressive sub-types within the more mixed manic-depressive or ' surgent ' group.

instead of tests was not a legitimate procedure, and was likely to be 'both futile and misleading' (p. 178). Accordingly, to test the differences between us, it was suggested that, with the aid of our joint research students, the problems chiefly affected should be worked over afresh by both procedures.[1] Their investigations, we hope, will shortly be published.[2] Meanwhile, Stephenson himself has given a preliminary account of the highly interesting results obtained by his modified technique in several short and illustrative inquiries. So far, however, as I have already indicated, I fail to see any serious discrepancy between either our figures or our deductions.

Here we may confine ourselves chiefly to those studies that bear

[1] At first it was suggested that the data already obtained with the various tests of *p* and *f* should themselves be correlated by persons. Stephenson, however, held that such a procedure would be invalidated by differences in the unit of measurement: 'since each of our tests has a different unit, they cannot be correlated by columns and the factors would be meaningless.' This did not appear to me to be fatal, since, if they were first expressed in standard measure for tests, the original measurements could still be correlated for persons.

A more serious objection, as it seemed to me, was that the data supplied by the tests, being confined to *p* and *f*, did not cover a sufficiently wide field of human behaviour: in correlating tests, our persons must be a fair sample of the population of persons; and similarly, in correlating persons, our traits must be a fair sample of the population of traits. Accordingly, for a valid use of the new method it appeared essential to take, not a number of tests for just a pair of important qualities, but a more comprehensive list of traits (such as McDougall's catalogue of the primary emotions) professing to represent the entire emotional life. In his later studies Stephenson has, I think, tacitly accepted this point of view, since the list that he has used, though drawn from Kretschmer's psychobiogram instead of McDougall's catalogue, includes (with a slight change of nomenclature) practically every trait in my own list except sex. He himself, however, lays stress on a somewhat different reason, namely, not so much to make the list a better sample of the total universe of individual actions, but rather to reduce the probable error by increasing the number of items over which the correlation is carried. Accordingly, since number is his main aim, he does not object to repeating estimates of the same trait in slightly different forms—a procedure I should be tempted to criticize on the ground of the irregular distribution that the sample of traits would then exhibit.

[2] Some of their earlier results are described by Stephenson [97]. Fuller accounts will be found in their M.A. and Ph.D. theses (University of London Library).

more particularly on the 'inhibited' and 'uninhibited' types of temperament, since it is mainly to my earlier statements about these types that Stephenson takes exception. His first study ([97], pp. 357–9) is most closely comparable with my own, since he adopts much the same list of traits, applies my rating scale for temperamental qualities, and follows the summation procedure outlined in my memorandum: a general factor is first eliminated; negative signs are reflected; a bipolar factor is then extracted (instead of two positive factors); and the factors are given a psychological interpretation as they stand: for we both agree that in such cases the 'rotation of axes,' insisted on by Thurstone, is here unnecessary. The data were self-ratings from 18 students in one of our seminar-classes: as the observer was different in each case—namely, each student himself—persons only can be correlated. After removing the general factor (which merely indicates the tendency of the self-observers to "estimate themselves highly for desirable qualities") the second factor to emerge has negative as well as positive saturations and *only one zero*. It is described as "representing the cyclothymic-schizothymic types." Here, therefore, Stephenson's chief type-factor is *almost exactly identical* with the bipolar second factor found in my own researches.

As we have seen, however, Stephenson would in general prefer, not to eliminate the general factor, but to use the group-factor method throughout instead, so as to obtain two positive factors instead of one bipolar factor. Where the data are derived from sharply discontinuous groups, as in his second investigation, I should, of course, agree that this was the correct procedure: but even here the final outcome appears to be the same. In his second investigation ([98], pp. 363 f.) he takes a composite sample containing 6 normal, 5 manic-depressive, and 5 schizophrenic persons, the latter being patients already tested at Horton Mental Hospital in the previous research on p and f. The same observer now estimated all the persons; consequently correlations can be legitimately calculated for traits as well as for persons, and I hope will ultimately be published for comparison. On analysing the correlations between persons, two factors are found—one for the 'manic-depressive type' (pathological extraversion) and the other for the 'schizophrene type' (pathological introversion), the second being virtually the inverse of the first. But these are precisely the 'clinical types' whose characteristics corresponded with the results of testing for p and f. Thus, the 'new factors' obtained by correlating persons point to exactly the same classification as the old factors obtainable by correlating tests or traits.

The identifications which I have proposed could be proved or disproved more conclusively if we had fuller information about the nature of the factors discovered. As a rule, each is just given a single descriptive name. In none of the experiments so far mentioned is any quantitative specification of the factors attempted, e.g. by computing factor-measurements for the several traits (which are, of course, equally trait-descriptions for the several factors). After all, to make up a group containing (*a*) manic-depressives and (*b*) schizophrenes, and then to discover by factor-analysis that the group comprises two antithetical types, namely, the manic-depressives and the schizophrenes, yields of itself no great addition to our theoretical knowledge—though no doubt it may be of great practical interest as confirming the diagnosis of the patients, provided of course the doctors making the diagnosis did not also make the assessments. The problem for the psychologist is to ascertain what are the traits that more particularly characterize the schizophrene and the manic-depressive respectively. Are we, for example, right in equating them with the 'schizothymic-cyclothymic types' of his preceding research? Or again what degree of relative predominance do the several Kretschmer traits exhibit in each? May we sum them up by saying that they show the inhibitive or the aggressive emotions developed to excess?

In a more recent research,[1] Stephenson carries his analysis one stage further: he applies the regression-equation, deduces factor-measurements for the traits, and so provides material for an answer to these questions. If what he terms my 'reciprocity principle' is correct, these factor-measurements should correspond, not only with my own factor-measurements for traits obtained by correlating persons, but also with the saturations for traits obtained by correlating the traits themselves. If, on the other hand (as he declares), the principle "involves such arbitrary foundations that it scarcely merits a moment's consideration," then there should be no resemblance whatever. He bases his figures on correlations between 21 normal persons, calculated from self-estimates for a modified list of 22 traits. He obtains a bipolar table of correlations, and deduces two antithetical types and two contrasted factors. Once again, the second factor is virtually the same as the first with the signs reversed: indeed, if we eliminate what I regard as the irrelevant 'general factor,' the correlation between the two sets of saturation coefficients rises to practically $-1{\cdot}00$.

On turning to his list of factor-measurements we find that

[1] *Character and Personality*, IV, iv, pp. 303–4.

his first factor and type are characterized by "consistent flow of energy," and are "pushing, tenacious, headstrong," "fluent in speech," and almost entirely free from "inhibition" or "obstruction"; the second factor and type have "inhibition" and "obstruction" as their most well-marked traits, and are "shut in and restricted." Surely the factors thus specified are almost exactly the same as those underlying my own earlier distinction between the uninhibited (aggressive or sthenic) type and the inhibited (obstructed or asthenic) type respectively. What is more, they surely have a striking resemblance to the factors previously discovered and described by Miss Studman and himself by correlating tests, and identified by Miss Studman with my own. However much, therefore, we may differ about procedure and details, we seem after all to be in almost complete agreement about the two main types that so constantly recur, and about their essential characteristics.[1]

[1] Unfortunately, owing largely to practical difficulties beyond the investigators' control, in none of the experiments so far carried out has it been possible to observe all the conditions that I suggested as necessary for a fair comparison. These were "(i) that the group of persons used in correlating persons should be the same as those used in correlating traits; (ii) that the set of traits should be the same; (iii) that both persons and traits should be selected from their respective universes on an assignable basis, so as to form (e.g.) an approximately normal, homogeneous, or rectilinear sample of the 'population'; (iv) that the same observer should make, or at any rate standardize, the assessments for every person in the group." In the experiment with hospital patients the observer was the same, but the trait-assessments could not be normally or evenly distributed, because the group was composed exclusively of extreme cases; in the two experiments with students, the trait-assessments might no doubt have been fitted to the normal scale employed in other inquiries, but unfortunately the observers were different: here, therefore, Stephenson has followed the same procedure as Dewar and assumed that the 'traits' are normally distributed *within the same person*: (this hardly seems legitimate with a list largely made up of antithetical pairs; Dewar deliberately selected her 'traits'—i.e. test-pictures —so that they might reasonably be regarded as constituting a normal distribution). These shortcomings are, of course, almost inevitable in preliminary studies: as my attempts can testify, it is always very difficult to find any *one* observer who can assess every student in a large group with the same accuracy and detail as the medical officer in a mental hospital can assess all his patients.

In the last of the illustrative experiments described by Stephenson—'experiment No. 4: complementary R and Q analysis' (carried out by one of his students)—the difficulty was overcome by using test-measurements for both types of analysis. The original intention was to

Since the above paragraphs were written, Stephenson has published a further article [1] of great interest, in which he discusses somewhat more fully the factorial analysis of temperamental types by means of correlations between persons. The main contention of my original paper [30]—that the chief temperamental types, as described in non-statistical literature, can be verified statistically and interpreted in terms of factors revealed by correlational analysis—he accepts in principle: but he insists (1) that the correlations must be correlations between persons, not between traits, and (2) that, if types are to be established, these correlations must be analysed by Q-technique, and by Q-technique alone. To show how his method differs from that described in my last article, he takes a set of temperamental assessments for 34 persons and analyses the data in two alternative ways—first, by the two factor-theorems of Q-technique and secondly by a multiple-factor technique, which he takes to be equivalent to my own.[2] The argument that a type is a " pattern of tendencies " and, as such, is capable of unidimensional measurement, he emphatically rejects. Moreover, he claims that the temperamental types he discovers are ' limited

correlate the rows and the columns of one and the same complete matrix: but the labour of calculating correlations for 100 persons seemed prohibitive; so only a selection was actually used for ' Q-technique.' Even so, the correlation tables were presumably too large to print, and, indeed, no figures at all are cited; but the investigator's own conclusions evidently imply (as Stephenson's quotations indicate) that in this experiment the two sets of factors were virtually, if not precisely, the same ([97], p. 360). In his thesis cited above—*A Statistical Study of Physical and Mental Types* (University of London Library)—the problem is discussed more fully, with an illustrative calculation from physical measurements: it is found that " the same results are obtained either by analysing correlations between persons or by analysing correlations between traits." For the first bipolar factor " a correlation of ·90 was obtained between the factor-measurements derived by factorizing traits and the factor-loadings or saturations derived by factorizing persons " (*loc. cit.*, pp. 213, 221).

[1] ' A Methodological Consideration of Jung's Typology,' *J. Mental Science*, LXXXV, 1938, pp. 185–205. A more adequate answer to his criticisms will, I hope, shortly be published in the same journal.

[2] Actually the procedure does not conform to mine; but that is doubtless because, in my published articles, the abridged description of my method was not altogether clear. His own procedure is to begin by selecting 4 or 5 persons representing each of his main types. Naturally, therefore, his correlations reveal factors corresponding to these types, and his frequency-distributions show discontinuous groups. If, however, we are allowed to select our representatives in advance, surely we could demonstrate any types we liked!

types,' i.e. they are not (as I had maintained) end-sections of a continuous, symmetrical, and nearly normal distribution, but sharply demarcated categories or groups. Frequency distributions, however, he does not give, and it would seem that his conclusion is based rather on surmise than on any actual survey of a representative population.

Space will not allow me to reply to these further criticisms here. Most of his points, I fancy, have been covered by the fuller exposition given in the earlier pages of this book. It will, therefore, be sufficient to point out that, once again, our minor divergences on points of technique seem quite outweighed by the similarity of our results. Except for slight differences in nomenclature, his concrete conclusions really corroborate my own. Thus, he explicitly recognizes—(i) a " factor of general emotionality," " something akin to *emotional instability*, uncontrolled and innate " ; (ii) a more specialized factor accounting for extraversion, " related to inadequate *w* " which he apparently identifies with ' inhibition ' ; (iii) several minor factors, two of which seem to turn on the same contrast as my distinction between the 'intuitive' and 'analytic' types, while a third seems analogous to my ' sensory ' factor. He himself treats these various types as positive and co-ordinate groupings, all on the same level of classification ; whereas my procedure exhibits them as bipolar antitheses, providing a succession of cross-classifications, and so yielding one subdivision within another, each classification following in order according to the amount contributed to the total variance by the corresponding factor. But these are formal rather than material differences. They do not affect the essential outcome : namely, that *nearly all the factors discovered by factorizing correlations between persons with Q-technique turn out after all to be much the same as the factors originally demonstrated by factorizing correlations between traits and subsequently confirmed—at any rate to a large extent—by factorizing correlations between persons by so-called P-technique.*

Spearman's View of Types.—Seeing that the controversy began with a discussion of " Spearman factors in psychiatry," let us glance in conclusion at Professor Spearman's own views, as expressed in his monumental work [113], which has appeared since the foregoing was written. Spearman clearly holds that the analysis of correlations between *traits* (as distinct from correlations between *persons*) would be sufficient to reveal the existence of mental types, *if* they existed. In his earlier work

Since the above paragraphs were written, Stephenson has published a further article[1] of great interest, in which he discusses somewhat more fully the factorial analysis of temperamental types by means of correlations between persons. The main contention of my original paper [30]—that the chief temperamental types, as described in non-statistical literature, can be verified statistically and interpreted in terms of factors revealed by correlational analysis—he accepts in principle: but he insists (1) that the correlations must be correlations between persons, not between traits, and (2) that, if types are to be established, these correlations must be analysed by Q-technique, and by Q-technique alone. To show how his method differs from that described in my last article, he takes a set of temperamental assessments for 34 persons and analyses the data in two alternative ways—first, by the two factor-theorems of Q-technique and secondly by a multiple-factor technique, which he takes to be equivalent to my own.[2] The argument that a type is a " pattern of tendencies " and, as such, is capable of unidimensional measurement, he emphatically rejects. Moreover, he claims that the temperamental types he discovers are ' limited

correlate the rows and the columns of one and the same complete matrix: but the labour of calculating correlations for 100 persons seemed prohibitive; so only a selection was actually used for ' Q-technique.' Even so, the correlation tables were presumably too large to print, and, indeed, no figures at all are cited; but the investigator's own conclusions evidently imply (as Stephenson's quotations indicate) that in this experiment the two sets of factors were virtually, if not precisely, the same ([97], p. 360). In his thesis cited above—*A Statistical Study of Physical and Mental Types* (University of London Library)—the problem is discussed more fully, with an illustrative calculation from physical measurements: it is found that " the same results are obtained either by analysing correlations between persons or by analysing correlations between traits." For the first bipolar factor " a correlation of ·90 was obtained between the factor-measurements derived by factorizing traits and the factor-loadings or saturations derived by factorizing persons " (*loc. cit.*, pp. 213, 221).

[1] ' A Methodological Consideration of Jung's Typology,' *J. Mental Science*, LXXXV, 1938, pp. 185–205. A more adequate answer to his criticisms will, I hope, shortly be published in the same journal.

[2] Actually the procedure does not conform to mine; but that is doubtless because, in my published articles, the abridged description of my method was not altogether clear. His own procedure is to begin by selecting 4 or 5 persons representing each of his main types. Naturally, therefore, his correlations reveal factors corresponding to these types, and his frequency-distributions show discontinuous groups. If, however, we are allowed to select our representatives in advance, surely we could demonstrate any types we liked!

types,' i.e. they are not (as I had maintained) end-sections of a continuous, symmetrical, and nearly normal distribution, but sharply demarcated categories or groups. Frequency distributions, however, he does not give, and it would seem that his conclusion is based rather on surmise than on any actual survey of a representative population.

Space will not allow me to reply to these further criticisms here. Most of his points, I fancy, have been covered by the fuller exposition given in the earlier pages of this book. It will, therefore, be sufficient to point out that, once again, our minor divergences on points of technique seem quite outweighed by the similarity of our results. Except for slight differences in nomenclature, his concrete conclusions really corroborate my own. Thus, he explicitly recognizes—(i) a " factor of general emotionality," " something akin to *emotional instability*, uncontrolled and innate " ; (ii) a more specialized factor accounting for extraversion, " related to inadequate *w* " which he apparently identifies with ' inhibition ' ; (iii) several minor factors, two of which seem to turn on the same contrast as my distinction between the 'intuitive' and 'analytic' types, while a third seems analogous to my ' sensory ' factor. He himself treats these various types as positive and co-ordinate groupings, all on the same level of classification ; whereas my procedure exhibits them as bipolar antitheses, providing a succession of cross-classifications, and so yielding one subdivision within another, each classification following in order according to the amount contributed to the total variance by the corresponding factor. But these are formal rather than material differences. They do not affect the essential outcome : namely, that *nearly all the factors discovered by factorizing correlations between persons with Q-technique turn out after all to be much the same as the factors originally demonstrated by factorizing correlations between traits and subsequently confirmed—at any rate to a large extent—by factorizing correlations between persons by so-called P-technique.*

Spearman's View of Types.—Seeing that the controversy began with a discussion of " Spearman factors in psychiatry," let us glance in conclusion at Professor Spearman's own views, as expressed in his monumental work [113], which has appeared since the foregoing was written. Spearman clearly holds that the analysis of correlations between *traits* (as distinct from correlations between *persons*) would be sufficient to reveal the existence of mental types, *if* they existed. In his earlier work

he found little or no clear evidence for 'group-factors,' such as would confirm the existence of such types, either on the intellectual or on the temperamental ('orectic') side; accordingly, he suggested that all the differences between the so-called extraverted and introverted types may be explained by differences in some single factor like *p* (in the sense of 'perseveration'), and argued that, when once such a basis has been established by correlating behaviour-tendencies, "all of what has been said about such types may possibly be nothing but its natural consequences."[1] This view he still maintains with an increasing array of evidence; and, in his final discussion of the problem, he concludes that nothing is left of temperamental or 'orectic' types except the 'orectic factors' obtained by such correlations.[2]

The General Factor in Correlations between Persons.—These more theoretical disputes, however, about the identity of the factors got by correlating persons and correlating traits respectively we may now leave on one side, and keep mainly to the results of correlating persons. Here, whatever be our view upon the broader issue, we can at any rate agree upon the narrower and more practical point; namely, that, whether we calculate a saturation coefficient for each person by an elaborate factor-analysis (as in Tables I and V) or whether we simply correlate his empirical marks or measurements with a key set representing the hypothetical types (as in Table III), the figures obtained will be virtually the same, or at any rate sufficiently close to be treated for all practical purposes as equivalent. This being so, may we not now rely, for the needs of everyday assessment, on the simpler method of direct correlation with the 'key personality'—i.e. with the theoretical 'standard person'?

Before consenting to do so, there is one passing question that the practical worker is very prone to put. If correlating persons affords such a simple way of measuring their approximation to this

[1] Cf. *Abilities of Man*, chap. iv, esp. pp. 52 *et seq.*, and pp. 82, 305.

[2] *Psychology Down the Ages*, Vol. II, chap. xlii, § 7, 'Fate of Orectic Types': cf. chap. xxxviii, 'The New Typology.'

type or to that, " how is it," as Stephenson himself inquires, " that it has been neglected for so long ? " ([98], p. 366). The answer is, in the first place, that it has not been neglected so completely as is commonly supposed. Davies, Moore, and myself used it for assessing imagery-types as long ago as 1912 ; for determining types of character, temperament, æsthetic appreciation, and the like, it has often been employed in practical work ; [1] and in a recent survey of the literature, over forty investigations are cited in which the method of correlating persons had been used [130]. Nevertheless, it must be admitted, the few theoretical investigators who have relied upon it seem, as a rule, to have done so with reluctance and apologies. For this the chief reasons have already been stated. But there is one consideration that is especially pertinent here. When we correlate persons, we are apt to introduce an over-shadowing general factor, eliminated in correlating traits, which may be quite irrelevant to our main issue. In certain inquiries, as we have seen, this general factor may of itself provide the centre of interest ; but in researches on type-psychology, such as the present, it is an unnecessary and obscuring factor, and, what is worse, even a fallacious factor—a kind of halo effect.

This has not always been recognized. But the effect can easily be seen if a large number of judges are invited to grade or mark the same person or persons for a list of traits such as that used in my tables here. Some years ago, in *The Measurement of Human Capacities* [2] I suggested, as an ' exercise in practical diagnosis,' that readers should grade four delinquents (whose portraits were repro-

[1] It was included, for example, in the summary of test-methods prepared for the Board of Education in 1924 (Burt, *ap. Report on Psychological Tests of Educable Capacity*, pp. 58–9). In its graphic form the principle has been very freely used for constructing, either mentally or explicitly, a characteristic ' profile ' to represent the ideal type or ' standard personality,' and noting how closely each individual's ' profile ' approximates to it.

[2] Oliver and Boyd, 1927. Year by year I have tried the same experiment with my classes, sometimes at the beginning, sometimes at the end of the course. In addition, each student is asked to rank himself and at least one other person well known to him and to the group. In using portraits, my object has been not so much to prove or disprove the value of physiognomy as a guide to character, as to bring out the uses and limitations of impressionistic judgments, when systematized by a scheme such as the one we have been using here. When the persons assessed are the assessors themselves, the general factor for persons is still further coloured by the relative desirability of each trait, which, of course, is much the same for all : no one, for example, cares to call himself ' flighty ' or ' fanatical ' ; hence in Stephenson's lists these qualities sink towards the lower end of the scale for *both* his antithetical

duced) in accordance with a printed schedule of intellectual and temperamental characteristics. Over 200 correspondents have sent me their gradings. For one and the same child their assessments correlate with each other to an average amount of ·67. But does that mean that they agree to anything like that extent with the true order for that child? Omit the photograph; ask the same judges to rank the qualities specified in order of strength for any person they like to take. Though each chooses a different person, there will still be a positive correlation between the orders, ranging from about ·20 to ·45! This implies a *general* order towards which every human being is tacitly assumed to tend, or at any rate everyone belonging to the same age, sex, race, and social class. We all approximate more or less towards this general pattern. In all of us anger tends to be a stronger passion than curiosity, and fear than submissiveness or disgust; just as in all of us, including the extreme 'pyknic,' height remains greater than breadth, and breadth greater than thickness: even the fat boy in Pickwick was "not quite so big round as he was tall."

Those individuals who correlate most closely with this general order will be popularly regarded as the most human. Actually, of course, such an order is more likely to reflect opinion than fact, the unformulated psychology of the judges rather than the actual disposition of the group that is judged. Indeed, in some cases it may be largely an artefact, depending, for example, on such accidental conditions as the extent to which each trait is accessible to observation. Thus, cautious observers will tend to rank sex low in all persons, because they 'know little or nothing of that side' of the persons to be judged; on the other hand, the Freudian will usually rank sex high in conformity with the doctrine he accepts. In Table VI, under the heading 'p.i.,' I give the average saturation coefficients for the first or general factor for persons, derived from all the rankings I have obtained. It will probably be agreed that it represents the popular notion of the relative strength of the motives named as ingredients of our common human nature.[1]

types: ([92], p. 302; cf. Guilford, *Psychometric Methods*, pp. 276–7). Similarly, in my own list unobjectionable or pleasing traits (e.g. curiosity, tenderness) and pleasant emotions generally (e.g. joy, self-assertiveness) are, on the average, ranked much higher in self-ratings than unpleasant emotions (e.g. sorrow, disgust) or traits that provoke criticism (e.g. anger, sex).

[1] When the popular notion is obtained by getting people to rank the emotions in order of strength *without* reference to individual persons, sex is placed much higher. In rankings given for males by young males (e.g. students) it rises to the top. In the printed table I have not thought it

These various results show clearly the dangers of 'direct judgment' (cf. [114], pp. 163–4), and the risks of accepting a simple ranking of traits as revealing the special characteristics of the particular person judged. When our concern is primarily with temperamental types, we can, and I think always should, *eliminate this general factor* either by partial correlation or, more simply, by requiring our judges to mark each trait, not for its comparative strength in the individual viewed in isolation, but rather for its variation above or below the average or normal in the entire group or population from which he has been picked. That, however, is a task which can be performed with accuracy only by a trained and experienced judge who has already had wide opportunities for observing the kind of persons to be assessed. Nevertheless, under specifiable conditions the 'general factor' introduced by simple ranking remains constant enough for us to allow for it, at least in some crude measure: (a simple method will be described in the next chapter, p. 426).

necessary to tabulate the average coefficients for the two sexes separately. For women, tenderness, submissiveness, fear, curiosity and disgust have a higher average rank than for men, while sex and anger (pugnacity) have a lower: this again tallies with the common notion.

CHAPTER XIX

FINAL CONCLUSIONS

A. Measurement of Types

The foregoing considerations show that, if we propose to use correlation or factor-analysis for a scientific determination of temperamental tendencies, there are several preliminary problems to be decided at the outset. Whether we carry out a factor-analysis in full, or are content to calculate the required correlation more simply and directly, we are, in effect, correlating (i) measurements for the type, which we take as a permanent standard of comparison, with (ii) measurements for the particular person observed which have been obtained *ad hoc*. Suppose, then, we are contemplating a survey of temperamental tendencies among the general population: two practical questions must be answered first of all: (i) What special set of measurements are we to take for the key personality or type? (ii) How are we to secure individual measurements so that they shall be fairly comparable with those for the type?

(i) For the permanent standard, figures obtained from a single batch, even if far larger and more representative than the group here used to illustrate the modified procedure, could hardly be accepted without much wider confirmation. The methods, however, first described in my paper of 1915, have been applied to a number of miscellaneous groups; and to the data originally summarized a good deal of further material has since been added. From each of the larger groups so studied I have calculated a set of factor-measurements for traits, mainly by the abridged procedure here described,[1] and have then averaged

[1] With the earliest groups the figures were obtained by correlating traits, applying the summation method, and determining the factor-measurements for persons by the usual method of regression. The results have since been checked both by the least-squares method and by correlating representative

the figures thus obtained. These averages are shown in Table V (columns headed ' p. ii ' and ' p. iii '). Here we shall be concerned with the second factor only ; and for this the figures in the column headed ' p. ii ' will be taken as our standard of comparison. A large positive correlation with these figures will indicate a sthenic or extraverted temperament ; a large negative, an asthenic or introverted.

A glance at the differences between the final weights for different traits shows that they approximate, if anything, more closely to a rectilinear series than to a normal distribution (cf. Fig. 1, line e—e). Hence, a plain ranking (such as that shown to the right of the decimal coefficients in the table) might seem sufficient. Could we trust a similar ranking for the particular person to be assessed, we should have only to find and sum the rank differences. The corrected correlation, if desired, could then be read off at once from a graph.

(ii) In obtaining comparable measurements for individuals, the chief difficulty, as we have just seen, is to eliminate the influence of the ' general factor for persons.' The simplest device is to treat the averages for the several traits from the very outset as equal. This means, as we have seen, that, wherever possible, we should obtain our initial assessments as deviations about the average manifestation of the particular trait assessed in the general population as a whole. Merely to rank the different traits themselves according to an order of relative strength, significance, or representativeness within each isolated person—though often the only available method in actual practice—is apt to be misleading, unless a correction is at

persons. Throughout, the factor-measurements for persons (obtained by correlating traits) and the saturation coefficients for persons (obtained by correlating persons) agree sufficiently well with each other, and with the direct type-correlations obtained by the shorter method, to justify the substitution of the latter for the bulk of the data. For aid in the earlier calculations, I am indebted to my former assistant, Miss V. Pelling, in the more recent to Miss G. Bruce and to students who have from time to time worked on small batches of material. In preparing the present chapter, I have been more especially indebted to Dr. A. J. Marshall, Research Assistant in the Department, for assistance with the final calculations here reported and for his kindness in reading through these pages.

once made either by partialling out the general factor for persons or by some equivalent device.

If a simple ranking is used both for the standard and for the individual examinee, the following device provides a crude but fairly appropriate way of eliminating the influence of the general factor. Such empirical rankings, when averaged, usually show a correlation of between ·2 and ·4 with the general factor for persons; and this in turn corresponds with a total rank-difference of about 30 out of a possible 60. Hence, for rapid assessments in the clinic, I suggest determining the coefficient for extraversion by the formula $r = 1 - \frac{\Sigma d}{30}$, instead of by the usual formula for correlation by rank-differences. This roughly allows for the general factor for persons; and, when gradings from inexperienced judges are alone available, the result as a rule seems to come fairly near to the coefficient that would be obtained with a more accurate grading, standardized in terms of the general population, and assigned by an experienced judge on the basis of actually observed behaviour.

For psychological surveys and for theoretical research a composite quantitative assessment is indispensable; but for practical purposes a simple synoptic chart or diagram, showing the original measurements plotted in a fixed order on a conventional scale, is a speedier and more informative device. Such graphs have been freely used both by teachers and by school psychologists for reviewing and card-indexing the educational attainments of individual pupils.[1] Similar diagrams, too, sometimes called 'mental profiles,' have been employed in vocational guidance, occupational analysis, and, indeed, most branches of individual psychology. And the same device would be of equal utility in clinical work.

The essential principles are illustrated in Fig. 1. The scale on the right is for the saturations describing the types; the scale on

[1] The general method is illustrated more fully by the 'psychograms' printed in the L.C.C. *Report on the Distribution and Relations of Educational Abilities* (figs. 9, i, ii, iii, and iv: these show graphs for the commoner educational 'types'—the 'verbal' and 'non-verbal' types, the 'non-arithmetical' type, etc.). There, as here, the traits are arranged in order according to the *secondary* factors rather than the general. The unit is the 'mental year,' which for educational abilities is almost exactly equal to the standard deviation during the middle of the school period; here the standard deviation has been employed.

I may add that the plotting of contour-graphs from correlations is a simple and neglected technique that may often take the place of the slower and more laborious computations that figure in factor-analysis. For example, besides

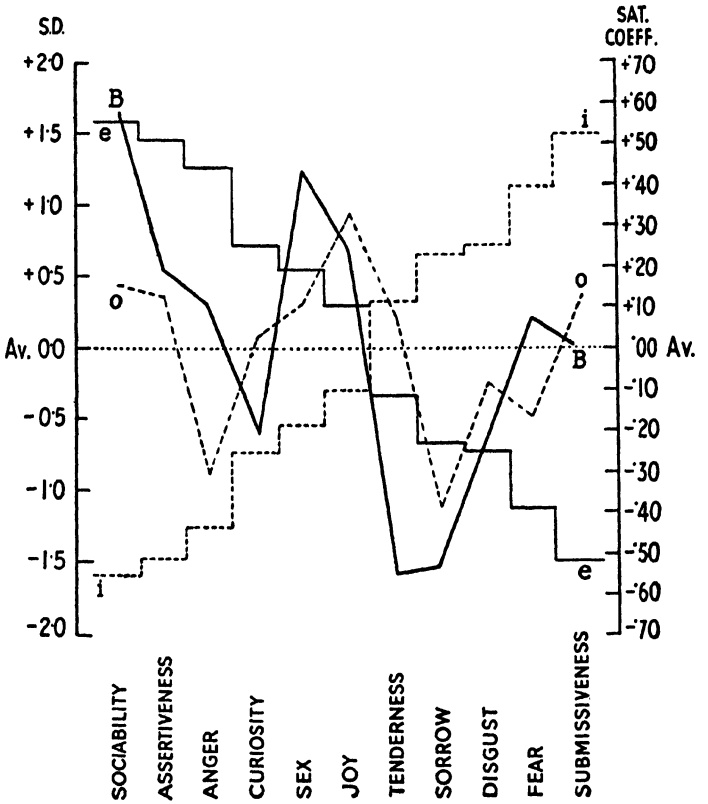


Fig. 1.—Temperamental Psychograms

e—e extravert type
i—i introvert type
o—o optimistic type
B—B 'profile' of person 'B'

the left is for the assessments reduced to standard measure. The thin continuous line (e—e) shows the 'profile' for the ideal 'extravert' or 'uninhibited type,' based on the figures given for p. ii, in Table V; the same contour turned upside down—the thin broken line (i—i)—gives the 'profile' for the ideal 'introvert.' These two estimates are plotted as column graphs, partly to show their step-like nature and partly to distinguish them from the others; they represent the basis of comparison when we are considering how far a child diverges towards the extraverted or the introverted type. The zigzag dotted line (o—o), shaped like a W, is based on the figures for p. iii, and represents the 'profile' for the ideal 'euphoric' or 'optimistic type.' The thick continuous zigzag line (B—B) gives the 'profile' of a single specimen individual (B in Table III), to illustrate the general method of construction and comparison. This represents the only line that need be drawn on the record-card itself (I use the backs of plain square-ruled record-cards: educational, vocational, or temperamental psychograms can be plotted on them as each case requires). It will be seen that the saturation coefficients allotted to student B for factors ii and iii really express the degree to which the contour of his graph (B—B) resembles the type-contours e—e and o—o respectively. But without any explicit calculation the practical worker quickly comes to recognize at a glance the three or four main points to be deduced from the individual curves. Thus, with temperamental assessments such as those for B in figure 1, (i) the general height of the line as a whole above the middle zero line represents the degree of the child's general emotional instability (unless already standardized at zero); (ii) the steepness and steadiness of the descent from left to right represents his approximation towards the extraverted type (correlation, ·81; an ascending line would represent an approximation to introversion); (iii) the approach towards the W-shaped contour of the dotted line indicates his tendency towards general cheerfulness (correlation, ·31; an inverted W would indicate the opposite); (iv) an outstanding peak represents an excess in one particular group of emotional or instinctive impulses (in this case sex), and an outstanding dip a defect. More specialized contours

plotting the column of *total* coefficients as above, it is highly instructive to plot the individual columns. Hierarchical order in the table then demands that all the contours shall resemble each other and the total contour: if not, the grouping of the resemblances may be used to locate group-factors. Thus, instead of calculating $\frac{1}{2}n\ (n-1)$ intercolumnar correlations we need only plot n columnar graphs and make the comparisons by eye (cf. [116], p. 346, footnote 7, and p. 300 above).

soon become familiar—e.g. those characteristic of the over-sexed child or the anxiety-neurotic.

B. Distribution of Types

We come now to our final problem—to examine the frequency-distribution of the temperamental tendencies as thus assessed within the population as a whole and within the chief samples of it that are conveniently accessible to character-assessment.

Writers who have announced type-distinctions like those between the extravert and the introvert, or the cyclothyme and the schizophrene, have commonly spoken as if such types must be mutually exclusive, or at any rate as if the amount of overlapping were all but negligible, so that nearly every person in the population could be allocated to one type or the other. Earlier psychologists indeed held such mutual exclusion to be an essential requisite of a sound logical classification. So far from believing that 'the commonest type is the mixed or intermediate type,' they quote with approval Kant's dictum: "Also gibt es keine zusammengetzten Temperamente."[1] In his recent discussions of the problem Dr. Stephenson supports this traditional view; and his main criticism against "writers from Professor Burt, on the one hand, to Clarke Hull, on the other" is that, on their assumptions, "types at most can only be regarded as extreme cases, tail ends of normal and continuously graded distributions." Any such deduction he emphatically rejects.[2] According to his own theory, "factors in Q-technique will usually not be universal: as Stern puts it, not everyone is a pickpocket; but we can determine *how many* are of one factor or type. Some persons would be of one type; some of another; some of neither . . . A person who has a zero saturation coefficient and correlates with the type factor by 0·00 will have none of that type in his make up" ([98], pp. 357–8). And in the last group (of 21 students) that he factorizes—though not in the first—there is a remarkable bunching of saturations for each of his two types first over the high values at the upper end of the scale and secondly over values in the neighbourhood of zero.[3] Such a bimodal distribution, I readily agree, is the very reverse of the frequency-diagram given in the paper of mine from which he quotes.

This notion of clear-cut types has been accepted by several psychiatrists; but it runs quite contrary to the current teaching

[1] *Anthropologie in Praktischer Hinsicht*, II, § 87.

[2] *Character and Personality*, IV, iv, pp. 295, 296. [3] *Ibid.*, pp. 298–9, 302.

of the academic psychologist. In the intellectual field the notion of discrete types has long ago been abandoned. By medical writers, it is true, one still occasionally finds the mentally defective described as "forming a special type apart to be sharply distinguished from the dull, who are merely unintelligent normals." But careful surveys by means of mental tests have shown—I think conclusively—that "the one group merges continuously into the other; there is no gap. . . . Apart from rare pathological cases, the mentally defective form simply the extreme tail of a continuous normal distribution." [1]

In the temperamental field similar surveys are urgently needed. The only extensive inquiry that I know is the oft-quoted investigation carried out over 30 years ago in Holland under Heymans and Wiersma.[2] Four hundred and fifty Dutch physicians were induced to make exhaustive reports on 2,523 individuals in accordance with a systematic questionnaire. The analysis of their data led to an eightfold classification of temperaments, based on three principles of division. Of these the most important was the contrast between high and low 'secondary function'—a perseverative tendency which was found to be high in persons of a *melancholic* type and low in persons of a *manic* type. The distribution was apparently bimodal: of the individuals studied less than 8 per cent. remained unassignable to one type or the other—not because they belonged to an intermediate category, but simply because in these few cases the available evidence left the true assignment doubtful. Suggestive as this research has proved, it would now, I imagine, carry little conviction as a study in the frequency-distribution of types. Its methods have been strongly criticized by Spearman; and its results can hardly be accepted as throwing genuine light upon our present inquiry.

The investigation of special cases and special problems, referred to me during my work as Psychologist for the London County Council, has incidentally yielded a vast accumulation of material that lends itself to analysis along the lines described in the foregoing pages. During thirty

[1] For the evidence, see my L.C.C. Reports on *The Distribution and Relations of Educational Abilities*, 1917, pp. 34 *et seq.*, and on *Mental and Scholastic Tests*, 1921, p. 163. Perhaps the most thorough criticism of the traditional view of types is that of Thorndike *Educational Psychology*, III (1923), chap. xvi: (temperamental types are not referred to as such, but they are tacitly covered by Thorndike's general conclusions).

[2] *Zeitschr. f. Angewandte Psychologie*, "Beiträge zur Spezieller Psychologie," I, 1908, pp. 313–83, *et seq.*

years' work in schools and other institutions I have collected quantitative assessments for many hundreds of persons, both normal and abnormal. At child-guidance clinics and mental hospitals other psychologists or physicians have doubtless done the same. My own assessments cannot pretend to be scientifically exact: but, in the hope of encouraging similar reviews, I venture to summarize them here for what they are worth.

The main groups with which I have been concerned are enumerated in Table VI. The children consist chiefly of boys and girls selected to form the large control groups used in my investigations of delinquency and backward children in the L.C.C. schools during 1913–31. With these are included a smaller group studied intensively for a research on vocational guidance, carried out under the Medical Research Council and the National Institute of Industrial Psychology: this partly accounts for the larger number in the group aged 10–14. Those under 14 are nearly all drawn from elementary schools. Those over 14 were either attending schools of the central or secondary type, or else had left school; since my own work was primarily school work, school pupils still form the majority even at these older ages; consequently this batch is probably of a more intellectual stamp than the general population. The adults were nearly all young men and women between the ages of 21 and 26. For them the data were collected mainly at two periods: in 1910–13 I was able to obtain assessments for a number of students at the University of Liverpool (the results obtained with this group were fully described in my British Association paper); from 1932 onwards I have been able to make further studies among students at the London Day Training College (now the Institute of Education) and later at University College. For non-academic adults I have collected data from two small clubs of working men and women formed at Settlements at Liverpool and London respectively, in which I happened to be residing.[1]

[1] I have very gratefully to acknowledge the assistance of colleagues at the settlements, of students at the colleges, and of teachers and others at the schools who assisted me.

TABLE VI

Distribution of Temperamental Tendencies

	Number in Group:	Correlation with Type-factor. Introversion ←																		→ Extraversion		Mean.	Standard Deviation.
		−·95	−·85	−·75	−·65	−·55	−·45	−·35	−·25	−·15	−·05	·05	·15	·25	·35	·45	·55	·65	·75	·85	·95		
Children :																							
Age 6–10 :																							
Boys . . .	78	0	1·3	0	2·6	3·8	2·6	5·1	5·1	7·7	6·4	5·1	7·7	7·7	10·3	7·7	8·9	5·1	6·4	2·6	3·8	·187	·432
Girls . . .	89	1·1	0	1·1	2·2	5·6	4·5	4·5	5·6	6·7	6·7	7·9	5·6	9·0	10·1	7·9	7·9	3·4	2·2	6·7	1·1	·128	·437
Age 10–14 :																							
Boys . . .	161	1·2	2·5	1·9	3·1	5·0	6·8	6·2	5·0	4·3	6·8	6·2	6·8	8·1	7·5	6·2	6·8	4·3	2·5	3·8	5·0	·082	·490
Girls . . .	176	0	1·1	2·3	4·0	5·1	6·3	5·7	6·3	6·8	7·4	6·3	7·9	9·0	8·5	7·4	6·3	3·4	2·8	2·3	1·1	·051	·430
Age 14–18 :																							
Boys . . .	92	2·2	5·4	7·6	4·4	6·5	4·4	7·6	5·4	6·5	5·4	6·5	4·4	7·6	7·6	6·5	4·4	1·1	3·3	1·1	2·2	—·090	·497
Girls . . .	123	1·6	0·8	2·4	3·3	4·1	4·9	5·7	4·9	5·7	5·7	4·9	5·7	6·5	7·3	8·2	7·3	6·5	4·9	5·7	4·1	·146	·496
Adults :																							
Students :																							
Men . . .	141	2·8	4·3	5·7	7·1	6·4	7·8	6·4	7·1	6·4	8·4	7·1	5·0	4·3	5·0	4·3	3·5	1·4	2·8	3·5	0·7	—·131	·483
Women . .	116	1·8	0·9	1·8	2·6	3·4	6·9	8·6	6·9	6·0	5·2	9·4	10·2	7·8	6·9	5·2	3·4	2·6	5·2	3·4	1·8	·045	·446
Working Class :																							
Men . . .	63	0	1·6	3·2	0	6·3	4·8	7·9	7·9	6·3	9·5	7·9	9·5	8·0	9·5	4·8	3·2	4·8	1·6	3·2	0	·025	·405
Women . .	57	0	1·8	1·8	1·8	3·5	5·3	3·5	5·3	10·5	7·0	7·0	10·5	7·0	8·8	7·0	5·3	3·5	5·3	1·8	3·5	·111	·431
Total . . .	1,096	1·2	2·0	2·8	3·5	5·0	5·8	6·2	5·9	6·4	6·8	6·7	7·1	7·5	7·8	6·5	5·8	3·6	3·5	3·5	2·4	·048	·471
Normal Distribution :																							
(*a*) *r*-method . .	100	[2·2]	1·4	2·0	2·8	3·8	4·9	5·9	7·0	7·8	8·3	8·5	8·2	7·7	6·9	5·9	4·8	3·7	2·8	2·0	[3·4]		
(*b*) *z*-method (skew) .	100	0·7	2·4	3·7	4·2	5·1	5·5	5·7	6·0	6·2	6·4	6·5	6·6	6·6	6·6	6·5	6·2	5·5	4·8	3·6	1·2		
(*c*) *z*-method (symmetrical) .	100	0·9	3·0	4·3	5·0	5·6	5·9	6·2	6·3	6·4	6·4	6·4	6·4	6·3	6·2	5·9	5·6	5·0	4·3	3·0	0·9		

TABLE VII

ANALYSIS OF VARIANCE

Source of Variation.	Sum of Squares.	Degrees of Freedom.	Mean Square (Variance).	Standard Deviation.
Between groups . .	9·9699	9	1·1078	1·0525
Within groups . .	232·5744	1086	0·2142	0·4628
Total	242·5443	1095	0·2215	0·4706

$$z = \log_e 1{\cdot}0525 - \log_e 0{\cdot}4628 = 0{\cdot}8216$$

With groups of the size indicated the 1 per cent. point of significance is given by $z = 0{\cdot}4437$.

Table VI gives the percentages within the main groups. In order to determine whether the variations in the means of the several groups are significant, a simple analysis of the variance has been carried out in Table VII. It will be seen that the variations are too large to be attributable merely to the fluctuations of random sampling.[1] So far as the general trend of the figures can be followed (Table VI), there would appear to be a fairly definite tendency for the amount of introversion to increase from the junior years to the middle of the school period, and again (at any rate among the boys) during adolescence. Before adolescence extraversion is apparently more prevalent among boys than among girls; but during adolescence and among adults introversion seems more frequent among the males—though here the groups may not be truly representative. At adolescence the variations are widest; and extreme types commoner than at any other period. Introversion, as might indeed be anticipated, seems more prevalent among the students than among the working men and women;[2] and, both among the students and among the working men and women, it seems slightly more frequent in the southern groups than in the northern, though here

[1] When the group-means are compared two at a time, a difference may be regarded as statistically significant if it exceeds ·13, or thereabouts.

[2] The workers are probably not a random sample of their class. Being taken from members of clubs, they probably include the more clubbable, i.e. the more extraverted, representatives.

the differences[1] are too slight to make it worth while to regroup the data according to locality.

In every group the distribution is unbroken: there is *no hint of any sharp demarcation between the two types*. The commonest cases are the intermediate cases; and the more salient instances of extraversion or of introversion arise simply as tail-ends of one continuous distribution. So far, therefore, the results definitely confirm the view put forward in my article as against those of its more recent critics.

Can we go on to claim that the general distribution is not only continuous, but approximately normal? Since the component groups are small and the differences comparatively slight, we may concentrate attention mainly on the totals (see Fig. 2). First, it is evident that the distribution of the frequencies is by no means perfectly symmetrical. For the whole group the average assessment (i.e. the average of the persons' correlations with the theoretical standard for the pure extraverted type) is + 0·048; the mode is at about + 0·3; zero, of course, would mark the ideal intermediate or well-balanced person. Thus, as the figures at the foot of the table plainly show, extraverts are more numerous than introverts. In part this may be due to the inclusion of particular age-groups or social groups among which extraversion is admittedly the more prominent characteristic (e.g. very young children who have not yet acquired full self-control, adolescent girls who have temporarily lost it, boys and girls from poorer homes where discipline or the lack of it conduces to impulsive rather than restrained behaviour). In part the skewness may indicate that the healthy, normal human being is naturally more inclined towards extraversion than the reverse.[2] But in part it may well imply that our standard of the

[1] Possibly 'racial' ([22], [129]). But the most marked of these differences were noticed in samples containing Jews and non-Jews. According to the assessments, the Jewish members would seem to form a decidedly abnormal group: they have therefore been excluded throughout.

[2] The use of the negative sign to mark the introvert does not mean that I regard introversion as a purely negative characteristic—due simply to the lack of a 'sthenic' or 'aggressive' factor. Indeed, as noted above, it is usually the 'inhibitive' or 'asthenic' factor that seems to indicate the addition of some more positive influence. The 'sthenic' type seems often a person whose high but natural emotionality develops freely and normally in the absence of repressing or inhibiting tendencies. On the other hand, the extreme introvert shows, far more frequently than the extreme extravert, the physical and mental signs of definite pathological disturbance. This is largely borne out by the comparative frequency with which

ideal or well-balanced personality, derived as it is from judges of a somewhat academic outlook (teachers and students), is too heavily weighted towards introversion, and is in need of further correction by a statistical adjustment such as that suggested above.

Secondly—and this is a more serious discrepancy—the frequencies at the extreme ends of the scale are unduly large: indeed, in the component groups the numbers having saturation coefficients of ·90 or over will seem astonishingly high to those who are familiar with the distributions of correlations obtained in the ordinary way. The standard deviation of the whole is 0·47 and of the component groups between 0·4 and 0·5, i.e. between a fifth and a quarter of the range; whereas, with a normal distribution of 1,000, we should expect it to be in the neighbourhood of a seventh, i.e. about 0·3. The histogram shows an unequal bulging on either side of its peak, which is itself pushed towards the right. There are no signs of a double peak. Yet it is scarcely surprising if, in earlier studies where small groups of only 15 or 20 have been used, so that the measurements are more thinly scattered, the investigators, noting how heavily the ends were weighted, have been led to infer that the true distribution is bimodal, and have consequently concluded that mental types, instead of being drawn from the opposite tail-ends of a single unimodal distribution, form in fact two discontinuous distributions, each with a peak or mode of its own. Nevertheless, it will not, I think, be difficult to show that both the anomalies I have mentioned arise merely from the methods by which the figures have been reached—partly from the method of estimation,[1] but mainly from the index of measurement employed.

'stigmata,' glandular and nervous disorders, and even relevant disease may be observed among adult introverts (see below, p. 439 and p. 385 above).

[1] The researches cited in support of a bimodal view were based on estimations supplied by students already familiar with the current descriptions of the types in question. Now, in all such judgments, as is well recognized, there is an unconscious bias, influencing nearly every judge, towards making the descriptions of each individual logically consistent. Such a bias will inevitably magnify the individual agreement with the ideal picture of the types, and so increase the high correlations at either end of the scale. The point is mentioned here because I believe my own data are by no means free from this suspicion. A number of the case-studies were carried out by research-students who were co-operating in the investigations on backward and delinquent children; from their previous psychological training these students were well acquainted with the supposed characteristics of the extraverted and the introverted types, and having decided that this child or that belonged to a particular type may have tended to fit their assessments accordingly.

The peculiarities of the index of measurement will be perceived more clearly if we compare the actual frequencies with the normal. From the ordinary table of the probability integral, we can determine precisely what frequencies would be required in theory by a normal distribution having the same mean and the same standard deviation as the present group. These theoretical percentages are printed in Table VI immediately beneath the percentages actually obtained. Since they have been calculated by taking the correlations just as they stand, I have labelled them '*r*-method.' Evidently the fit is far from good. In particular the theoretical distribution, reconstructed in this way, inevitably yields frequencies beyond the values of + 1·00 and − 1·00. But a coefficient of correlation, of course, cannot go beyond either of these limits. Hence, *if our index of measurement is to be expressed in the form of a correlation, true normality is precluded from the outset*; and we discover at once an obvious reason for the bunching at the tails.

We may, however, remove this limitation by a simple device. We may adopt for our conventional measure, not the correlation coefficient itself, but the number having this coefficient for its hyperbolic tangent. The range of the numbers so derived is unlimited in both directions; and, as is well known, their sampling distribution conforms almost exactly with the normal curve (cf. [110], p. 451 and refs.).

Let us then apply this conversion. We take for our adjusted measure [1]

$$z = \tanh^{-1} r = \tfrac{1}{2} \log_e \frac{1+r}{1-r} = r + \frac{r^3}{3} + \frac{r^5}{5} + \frac{r^7}{7} + \ldots$$

where r is the correlation (or saturation coefficient) and z the substituted measurement. In terms of z, the mean for the entire group is now ·06 and the standard deviation ·62. The size of the standard deviation is alone sufficient to dispose of any notion that the original assessments may have been assigned practically at random: for, in that case, with correlations based on $n = 11$

[1] I give various equivalent expressions, since different students may prefer to adopt different methods of making the transformation. Full tables of the hyperbolic tangent will be found in *Smithsonian Mathematical Tables: Hyperbolic Functions*, pp. 86 *et seq.*; or the calculation may be made from any sixpenny set of mathematical tables that includes 'natural, hyperbolic, or Napierian logarithms' or the 'exponential function.' Table VB in Fisher's *Statistical Methods for Research Workers* (p. 197) may also be used, if available: for such work as the present, however, this involves interpolation; hence we find it more convenient to employ a specially compiled table giving z values for r instead of *vice versa*.

assessments, we should have expected the standard deviation of z to be in the neighbourhood of $\frac{1}{\sqrt{n-3}} = \frac{1}{\sqrt{8}} = 0{\cdot}35$, whereas actually it is nearly twice that figure.

Accordingly, assuming a normal distribution, having the mean and the standard deviation specified above, we can now calculate what percentages of the whole group should fall between the successive values of r given in the table. These percentages are inserted in the last line but one. It is, however, almost equally interesting to inquire what frequencies we should expect with a normal distribution varying symmetrically about zero (the perfect temperament) taken as the mean. Accordingly, we may smooth away the asymmetry of the observed distribution by averaging the percentages on either side of zero and recalculate the standard deviation (it remains practically the same) and once more deduce the theoretical percentages fitting a normal curve. These further figures are appended in the last line of Table VI.

The effect of the hyperbolic transformation is vividly shown by plotting the frequencies above an r-scale and a z-scale respectively. With the r-scale, taking, that is, the correlations as they stand (Fig. 2), the distribution appears anything but normal: yet when expanded to a z-scale they show a reasonably close resemblance to a normal distribution (Fig. 3). When the theoretical percentages for a strictly normal distribution (Fig. 3, dotted line) are converted back to an r-scale (Fig. 2, dotted line) all resemblance to a normal curve is lost: and it is scarcely surprising if such flattened distributions have been supposed to put the hypothesis of normality entirely out of court.

The degree to which the various theoretical percentages fit the observed may be tested more precisely by the usual χ^2 method (i.e. dividing the squares of the discrepancies by the theoretical values and summing the ratios). The last of the hypotheses mentioned above—that the true distribution is symmetrical about the intermediate type (measured by zero)—yields, as was to be expected, the poorest fit of all: ($\chi^2 = 58{\cdot}09$; P—the probability that the divergences from the hypothetical distribution are due solely to errors of sampling—therefore lies, as Pearson's Tables[1] show, between 1 in 10,000 and 1 in 100,000). The figures based on the correlations as they stand are not much better ($\chi^2 = 54{\cdot}47$; P is therefore less than 1 in 1,000). The figures given by the

[1] *Tables for Statisticians and Biometricians*, Table XII ('Tables for Testing Goodness of Fit.' Enter the table with $n' =$ *one more than* the number of degrees of freedom: cf. [110], p. 418).

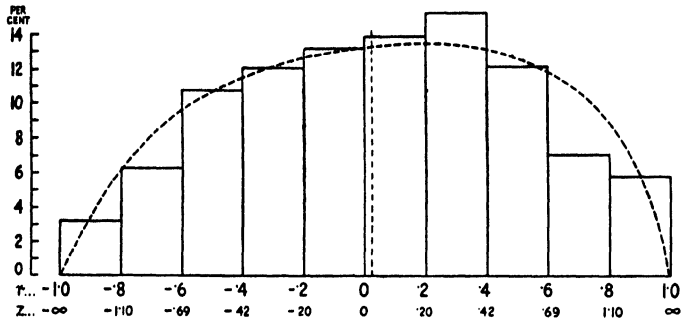


Fig. 2.—Distribution of Temperamental Types as measured by Saturation Coefficients (r).

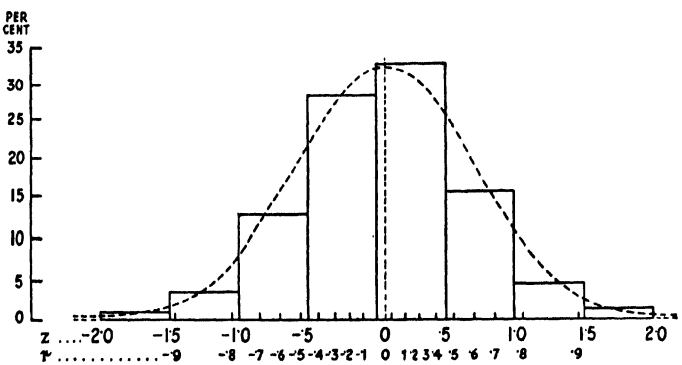


Fig. 3.—Distribution of Temperamental Types as measured by $z = \tanh^{-1} r$.

asymmetrical z-distribution fit much more closely ($\chi^2 = 31{\cdot}24$; $P = {\cdot}051$, i.e. just over 1 in 20).

It may be of interest to speculate what influences can reasonably account for the divergences that remain discernible. As the analysis of variance indicates, some of the additional spreading in the total figures is caused by the pooling of heterogeneous groups. But other causes are unquestionably at work.

In the component curves, as we have already noted, the most conspicuous deviations from normality consist in a heaping of frequencies near either tail end and to a less extent near the centre. Instances of the more extreme types, and instances of the well-balanced type, are more numerous than a perfect normal distribution would allow; and the excess appears most marked among the older persons. An examination of the individual case-histories suggests at least two explanations. First of all, every large and random group must inevitably contain several persons who present mild but definite pathological symptoms: one of the adult introverts, for example, was subsequently diagnosed as a case of dementia præcox; three of the extraverts were subsequently treated for hysteria; several others were noted as cases of suspected glandular disturbance (e.g. there were three instances of definite hyperthyroidism among the adult and adolescent extraverts and as many as 18 cases suggestive of mild hypothyroidism or mild hyperpituitarism among the adult introverts; 'high blood pressure' was reported in 6 cases, including 2 hyperthyroid).[1] Among the children, pathological cases were rarer: but at least 5 showed symptoms of mild chorea (4 girls aged 8–10, 1 boy aged 11). Secondly, among those exhibiting no discernible pathological symptoms, the more extreme temperaments seem often to have undergone a self-magnifying process, which has had a cumulative effect as the person has grown older. As a result of habit and of the repeated reactions of social intercourse, the introvert tends to become more strongly and more generally introverted, and the extravert, instead of being progressively suppressed, often develops an aggressive attitude that grows more and more defiant and more and more obstinately fixed. Possibly a similar tendency towards an all-round self-consistency operates in the moderate types: those who from the outset possess

[1] With a small group of 35 male students we found a correlation of ·54 between systolic blood pressure and extraversion, ·23 between extraversion and weight, and ·41 between systolic blood pressure and weight: (the latter figures are of interest as suggesting that correlation of the so-called cyclothyme temperament with pyknic physique may possibly be indirect, and due chiefly to the hypertension associated with excessive weight).

but little bias in either direction doubtless find it easier to conform to the stereotyped and conventional pattern of behaviour, and so, in later years, present personalities which, to the outward observer, appear unusually correct and well-balanced.[1]

We may, therefore, conclude that the distribution underlying the alleged antithetical 'types' is a continuous, unimodal, bipolar distribution, closely approximating to a normal distribution. But we must add that in any particular group the perfect symmetry of the normal curve is likely to be appreciably disturbed, partly by the inevitable inclusion of mild pathological cases and probably by other

[1] A footnote should perhaps be added dealing with one or two criticisms that have appeared since the above was written. Dr. Jameson asks whether the conclusions drawn in my previous paper [114] are not inconsistent with my own earlier acceptance of Mendelian principles as applicable to mental as well as to physical inheritance ([22], [23]): "do not these very laws lead us to expect sharply segregated types, 'objective' and 'subjective,' to use Burt's own designations, 'extravert' and 'introvert,' 'cyclothyme' and 'schizophrene' to use the more recent, and (in my view) more precise and informative terms?" Similarly, Dr. Stephenson in his latest paper on *Jung's Typology* argues, as we have seen (pp. 418–19 above), that the temperamental types distinguished by Kretschmer and Jung are to be regarded as 'definitely limited species,' not 'extreme ends of a normal frequency distribution': "mental types," he insists, "are as distinct as the genera and species of animals in the present epoch of fauna—as cats from dogs, and buttercups from daisies." In reply to Dr. Jameson's argument it may be said that, where a character appears relatively simple in its causal origin (e.g. to take the instances he quotes from my earlier paper, sex-characteristics or red-green colour-blindness), there we may expect sharply separated types: even in colour-blindness, however, experimental tests do not reveal distinctions by any means as clear cut as is popularly supposed. But where a character appears relatively complex in its causal origin (e.g. stature or temperament) there Mendelian principles themselves lead us to expect 'tendencies rather than clear-cut types' (cf. above, p. 246). In another paper I hope to show how the physical types popularly associated with distinctive temperamental qualities can themselves be measured by factorial means, and then proved to follow a continuous and approximately normal distribution. The 'types' that Stephenson cites from the plant and animal world are mutually sterile; hence there would seem to be no possibility of such types merging. But man and his 'types' are peculiar in two ways. First, man seems to have gone further than any other animal in producing markedly different varieties, which are nevertheless mutually fertile; moreover, in an intelligent and adaptable species like man a larger proportion of the more eccentric variations are able to survive and breed.

intrinsic influences as well. Thus, our final conclusions in regard to the distribution of temperamental tendencies are in complete accord with those previously reached for the distribution of intellectual capacities ([35], p. 34; [41], p. 162). Just as age, social environment, and special pathological influences disturb the symmetry and normality of the curves for intelligence, so they seem also to disturb the expected curves for the distribution of temperamental types. The deviations that result, however, are comparatively slight. For most practical purposes [1] (e.g. the construction of rating scales) it seems reasonable to assume that, except where special causes grossly intervene,

Secondly, owing to his strong migratory propensities, the different varieties thus arising have been continuously crossed and recrossed, so that any existing group (at any rate in civilized countries) exhibits a wider individual variability than any wild creature found in the same habitat. On Mendelian principles we should therefore expect to find, not the sharply segregated types and the bimodal distribution that Stephenson seems to anticipate in mixed communities of human beings as well as in mixed communities of cats and dogs, nor yet the more or less uniform blend that he seems to attribute to me, but countless recombinations, ranging from rare examples of one primitive type at one extreme to rare examples of the opposite type at the other, with most individuals varying continuously about a common mode and merely tending towards one extreme or the other.

However, I fully agree that it is high time that problems of mental inheritance and sex-differences should be taken up afresh by more up-to-date procedures. The psychological factorist will find it instructive to compare with his own procedure the statistical methods independently worked out by the geneticist for the detection and estimation of segregation, linkage, and heterogeneity in biological work: a convenient summary with references is available in 'The Measurement of Linkage in Heredity,' by K. Mather [126]. An up-to-date résumé of the literature on 'Methods of Assessing Temperament and Personality' is to be found in Dr. C. J. C. Earl's excellent chapter in *The Study of Society* (ed. F. C. Bartlett *et al.*). The concluding sentences of his section on 'Typology' (p. 234 *ad fin.*) seem to me to express, very clearly and succinctly, the right provisional conclusion.

[1] As I have argued elsewhere, the best general distribution for obtaining a close theoretical fit to the frequencies of mental traits is that given by the hypergeometric series. The series given by the symmetric binomial (from which the normal curve is derived) is merely a special case of this more general series. The latter distribution will include cases in which the contributory causes are *not* independent and in which the chances are *not* equal; and thus will enable us to introduce any requisite degree of asymmetry into the theoretical frequencies: (see *The Backward Child*, pp. 664 *et seq.*).

the distribution of innate temperamental tendencies may be treated as normal. Provisionally, and until more extensive surveys have been carried out, we might perhaps accept the converted z-distribution (last line of Table VI) as indicating the theoretical distribution of temperamental variations. If, however, the variations are measured by a coefficient of correlation (or its equivalent) then we must be prepared to accept distortions of the kind described above.

C. Summary

1. Reasons have been advanced for believing (*a*) that the temperamental factors deducible from assessments for groups of normal adults and children are essentially the same as those previously deduced from assessments for neurotic and delinquent children referred for clinical examination, and (*b*) that, whether persons or traits are correlated, the main type-factors (i.e. all significant factors except the first) remain virtually the same throughout.

2. Explicit calculation further shows that this equivalence of 'person-factors' and 'trait-factors' still holds good, even (*a*) when the correlations are derived from a relatively unselected group of normal persons, differing in general or average emotionality, instead of from a selected group, relatively homogeneous as regards this general factor; (*b*) when correlations are used instead of covariances; and (*c*) when the simple summation method is employed instead of the more laborious method of least squares. Criticisms of this so-called 'reciprocity principle' are shown to be inconsistent with the critic's own experimental results.

3. Explicit calculation verifies (what had previously been proved algebraically) that with the summation method the correlation of a person's measurements with the average measurements for the type is virtually identical with his saturation coefficient. It follows that, for practical purposes, in order to estimate the degree to which the several persons approximate to a given type, it is no longer necessary to calculate all the intercorrelations and then to perform a systematic factor-analysis: it is sufficient to correlate the

examinee's measurements with a standard series based on simple averaging and specifying the type.

4. Simple practical methods are suggested for recording either by a numerical coefficient or by a psychographic chart or 'profile' the essential temperamental characteristics and 'type' of individual persons.

5. By means of the methods described a large number of persons of both sexes and of different ages have been assessed for the bipolar factor which appears to underlie the difference between the so-called introverted and extraverted types. For each group examined the distribution proves to be not bimodal but continuous, the mixed or relatively well-balanced type being commonest of all. For the entire group the distribution is slightly asymmetrical, showing a small excess of extraverted cases. Within the component groups the amount of extraversion differs according to age, sex, and social or intellectual status, being greater among younger children and less in older or more intellectual males. In view of the age, history, and physical peculiarities of many of the introverts, it is suggested that introversion rather than extraversion is usually the positive characteristic, and may not infrequently be a late pathological development rather than a direct manifestation of inborn temperament. In any case, the 'factors' undoubtedly represent highly mixed and complex groups of causes.

6. For statistical purposes it is proposed to measure approximation to type, not by the saturation coefficient as it stands, but by its inverse hyperbolic tangent. With this transformation the frequency curve approximates closely to the normal. Thus the more typical introverts and extraverts appear in the main to be merely extreme cases taken from the opposite tail-ends of a normal or nearly normal distribution.

APPENDICES

APPENDIX I

WORKING METHODS FOR COMPUTERS

THE simple methods of computation which were originally developed for my work on mental and scholastic tests have since proved serviceable for many other problems; and it has more than once been suggested that a set of working instructions might be published in brief accessible form. The theory on which the methods rest has been formally set out elsewhere (e.g. [93]); but the experimental investigator is frequently unable to follow theoretical demonstrations in matrix algebra, and yet is desirous of attempting a practical analysis of his data. The following notes are abridged from longer roneo'd instructions, prepared some years ago for research students working at the London Day Training College, and revised from time to time in the light of their experience: these more detailed notes can be obtained on application to the Psychological Laboratory, University College, should fuller explanations be required.

The general procedure was first adopted[1] in order to meet the peculiarities of correlation tables that do not manifestly fit the simpler 'two-factor theory' of Spearman. As indicated in the original paper, the work consists essentially of three main steps. (i) With each of the tests, "the sum or average of its coefficients is taken as measuring its general tendency to correlate, and therefore as provisionally determining its position in the hierarchy." "The theoretical values," it was added, "can be obtained by various mathematical formulæ"; here I shall confine myself to two or three of the simplest. (ii) An "ideal hierarchy"[2] is then fitted to the observed coefficients by applying the "product equation" ($r_{ab} = r_{ag}r_{bg}$, where r_{ag} and r_{bg} denote the correlations of any two tests a and b with the common factor g). (iii) Finally, the theoretical correlations are subtracted from the observed correlations in order to study the "deviations" (or "residuals," as they

[1] 'Experimental Tests of General Intelligence,' *Brit. J. Psychol.*, III, 1909, pp. 94–177, esp. pp. 160–4, and Tables V and VI.

[2] It would have been technically more correct to say a 'matrix of rank one is fitted.' But the term 'hierarchy' is more familiar to the English student of psychology. It should be remembered, however, that the 'ideal hierarchy' as I have defined it (p. 149 above) is not the only type that has been described.

would now be called) for indications of further "subordinate factors." All through, however, the common underlying assumption is that any empirical table of correlations or covariances can be treated as the sum of a rapidly diminishing progression of hierarchies, each attributable to an independent factor.[1] At particular points and for particular methods certain supplementary assumptions have to be made, for which the chief justification is practical convenience. Where these diverge from the practice of other writers, or where they have been questioned or criticized, I shall take occasion to interpolate a note in their defence.

To determine the saturation coefficients r_{ag}, the easiest formula for general use is that first given and illustrated in my 1917 *Report on the Distribution and Relations of Educational Abilities*.[2] I have

[1] Cf. page 164.

[2] L.C.C. *Report*, No. 1868 (P. S. King & Son, 2s. 6d.), Tables XVIII–XXIV. It was more fully described in my 1915 paper on 'General and Specific Factors underlying the Primary Emotions' (briefly reported in [30]). Thurstone, who has since adopted the same formula, prefers to call it the 'centroid method' (see [84], 1935, chap. iii, eq. (13), p. 94): his working procedure, however, differs from mine (*a*) in the mode of estimating the missing diagonal values; (*b*) in the treatment of negative residuals; (*c*) in insisting on a subsequent rotation of the factorial axes. In discussing my introduction of the essential formula, Thomson adds: "it is not quite clear how he (Burt) would have filled in the blank diagonal cells" ([132], 1939, p. 25). How the missing value is filled in at the outset does not greatly matter, provided the figure is checked and corrected by the results thus provisionally obtained. With a nearly hierarchical table, the empty cells would have been filled by applying the proportionality formula given in my earlier article (cf. below, p. 450, footnote 1). With more irregular tables a direct calculation is no longer feasible; but, as I have elsewhere indicated, "the difficulty can be easily overcome by successive approximation: in the spaces for the self-correlations, trial values can be inserted by smoothing the several columns; these are then checked by computing $r_{kk} = r_{kg}^2$, when r_{kg} has been found by equation xxv" (i.e. by the summation formula, [93], p. 285: cf. 'Methods of Factor Analysis with and without Successive Approximation,' [102], p. 178). Where a multiplicity of factors is assumed, the check of course requires us to calculate $r_{kk} = \Sigma r_{kg}^2$, where g now denotes all the general and group-factors entering into test k: in either case the figure required is really the square of the multiple correlation of each test with its essential common factors, i.e. with the infinity of tests involving those same factors (p. 286). By 'smoothing' I meant the rough process that teachers and examiners so often adopt, and sometimes call by that name, when, for example, they wish to estimate an average or total for a boy who has been absent for one of a series of examination papers and so give him a rough allowance—an allowance which could no doubt be calculated more exactly, were such exactitude warranted (e.g. by a proportion based on the totals for all candidates except that boy and all papers except the paper missed, as in fitting a contingency table, p. 147); fortunately, as I pointed out, the figure suggested by the general trend of the correlation pattern is, as a rule, sufficiently near the mark to render much recalculation needless, unless a high degree of accuracy is required or the number of correlations to be summed or averaged is very small. For these latter cases I suggested a simple 'product formula' for estimating r_{kg} directly from the intercorrelations alone, viz.:

$$r_{kg} = \pm \frac{\left\{ \Pi\,(r_{kj}) \right\}^{\frac{1}{n-2}}}{\left\{ \Pi\,(r_{jj'}) \right\}^{\frac{1}{2(n-1)(n-2)}}} \qquad . \quad . \quad . \quad (j \neq j')$$

As is demonstrated in my fuller notes, in theory, (i) the process of successive approximation, if carried out mechanically and completely, with a matrix of rational coefficients, progressively reduces the reinserted diagonal elements (i.e. the values for the variances or communalities) to a set whose sum is the smallest possible that is compatible with the assumed positive-definite character of the completed correlation

termed it the 'summation formula,' to distinguish it from others that were tried at the same time, more particularly from the 'product formula,' which was based on the geometric instead of on the arithmetic mean. As in ordinary forms of averaging, so here, we may take either weighted or unweighted sums. The method of unweighted summation can be regarded as a simplification of, and a first approximation to, that of weighted summation.

I. SINGLE-FACTOR ANALYSIS

A. Simple-summation Method

Formula: Saturation coefficient $r_{ag} = \dfrac{\Sigma_i r_{ai}}{\sqrt{\Sigma_j \Sigma_i r_{ji}}}$.

The basic principle can best be understood if we begin with the simplest case of all, namely, that in which only one factor is involved, and in which, therefore, the correlations are assumed to form a perfect hierarchy (except perhaps for minor observational errors). Let us suppose that the observed correlations are those printed in the body of Table I below. The figures are fictitious, and have been artificially derived by multiplying each of the saturation coefficients (shown along the top and left-hand margins) by every other, in accordance with the product-equation. These saturation coefficients are presumed to be unknown; and our object is to rediscover them.

The working procedure is as follows:

1. Find the total intercorrelation of each test by adding each column of observed correlations: ($\cdot72 + \cdot63 + \ldots + \cdot45 = 2\cdot34$, etc.).

matrix (i.e. with a *real* factorial matrix); and (ii) such a set of diagonal elements yields (in general) a completed matrix of lowest possible rank. When weighted summation is used, the same completed matrix with the same rank and the same diagonal elements is ultimately reached. Hence, for the latter procedure, the diagonal values can first be calculated by simple summation so as to save labour.

In practice, however, I do not attach importance to the precise minimal rank as such. I find it difficult to conceive that an empirical correlation matrix, any more than an empirical covariance matrix, can have a definitely assignable minimal rank. Moreover, I should argue that the relative sizes of the variances ought not to be arbitrarily limited by intrinsic mathematical considerations. Indeed, it would seem defensible on theoretical grounds to assume slightly larger variances than the apparent minimum. These minor points make but little difference to the general procedure or to the results actually obtained: but they save the computer from laboriously struggling after an unwarranted precision in his final figures.

I may add that the more elaborate variants of the primary formula were originally devised for certain specialized problems arising out of my investigations for the London County Council, and were given (with other derivatives) in various *Reports*, usually without full proof. Here only those that are of frequent utility are included. The general line of proof has been indicated in the text. I am indebted to the Council for permission to incorporate material from their *Reports*.

2. Rearrange the whole table in the order indicated by these totals, beginning with the largest (2.34) and ending with the smallest (1·50). Since the correlation of each test with itself is assumed to be unknown, the leading diagonal is left empty.

3. Begin by making a provisional estimate of these unknown self-correlations. The simplest device is to consider by direct inspection what figures will best fit the general pattern. Thus, when the table has been arranged to form a descending hierarchy, we first insert in the space at the top of the first column a figure (·80, say) somewhat *larger*[1] than the largest intercorrelation in that column (·72); and then in the space at the bottom of the last column a figure (·25, say) somewhat *smaller* than the smallest intercorrelation in that column (·30)[2]; and finally fill in the intervening spaces down the central diagonal with figures descending by regular or irregular intervals corresponding with the regularity or irregularity of the successive intervals between the totals.[3] If these first trial estimates are likely to require correction, it will be better to write them, not in the body of the table, but below the totals, as has been done in Table I. Round figures (·80, ·65, ·50, . . ., ·25) will be sufficient for the first trial.

4. Add each estimated figure to the total of the corresponding column (2·34 + ·80 = 3·14, etc.).

5. Add these augmented or completed totals, to obtain their grand total (3·14 + 2·81 + . . . + 1·75 = 12·25).

6. Find the square root of this grand total ($\sqrt{12{\cdot}25} = 3{\cdot}50$).

7. Divide the augmented totals of each column by this square root (3·14 ÷ 3·50 = ·897, etc.). The quotients will yield a first approximation to the saturation coefficients required.

8. As a check on the arithmetic, add the estimated saturation coefficients. Their sum (3·50) should be equal to the square root already obtained in step 6.

9. As a check on the estimation, square the estimated saturation coefficients ($·897^2 = ·805$, $803^2 = ·645$, etc.). Their squares should

[1] The beginner, who does not trust his powers of smoothing the general pattern at sight, generally asks: how much larger? If a mechanical rule is desired, he can take the proportions between the two adjacent coefficients (or between the two column totals for the same tests) as a guide: e.g.

$$\frac{r_{11}}{·72} = \frac{·63}{·56} \text{ or } \frac{2·34 - ·72}{2·16 - ·72}, \text{ which gives } r_{11} = ·81.$$

[2] Thurstone, after considering half a dozen different procedures, recommends always filling in the gap with the highest correlation in the corresponding column ([84], pp. 89, 233). In the present case this would lead to a very erroneous result. In the first column the self-correlations should evidently be *larger* than the largest intercorrelation; in the last column (assuming there are no group-factors) it should evidently be, not as large as the largest, but *smaller than the smallest*.

[3] Here again, if the beginner requires a mechanical rule, it is the equality or inequality between the *differences* of the intervals that should be the guide.

be nearly identical with the trial figures originally proposed for insertion in the diagonal (·80, ·65, etc.).

10. If the agreement is not sufficiently close, repeat the calculations with new trial estimates. Take the squares of the saturation coefficients, first slightly readjusting them wherever necessary—e.g. if, like ·805, they are larger than the first estimate (·80), make them larger still (say ·81); if, like ·645, they are smaller, make them smaller still (say ·64). Add the adjusted squares to the original column totals (2·34 + ·81 = 3·15, etc.); and proceed as before. If a more exact estimate is required, the process of successive approximation must be continued, until figures of the required accuracy are obtained; but, after a little experience, one or two repetitions can be made to suffice.[1]

11. Calculate the hierarchy of theoretical correlations resulting from this factor by multiplying each of the saturation coefficients with the rest, as shown in the original construction of Table I (1st row: ·90 × ·80 = ·72, ·90 × ·70 = ·63, etc. 2nd row: ·80 × ·70 = ·56, etc.).

12. Subtract these theoretical values from the observed correlations, and test the significance of the deviations by the standard error of the observed coefficient or by χ^2, as suggested above (p. 339). Here, of course, the differences are zero.

B. Weighted Summation

Formula: Saturation coefficient $r_{ag} = \dfrac{\sum_i r_{ig} r_{ai}}{\sum_j \sum_i r_{ig} r_{ji} \div \sum_i r_{ig}}$

If the observed table is not an exact hierarchy, the method of weighted summation will produce a better fit; in fact, for a complete table (i.e. with known figures for the diagonal) it gives the best possible fit as judged by the method of least squares. With an ordinary set of correlations, where a unique determination of values for the diagonal is rather doubtful, the gain may seem scarcely worth the additional labour: but with a table of variances and covariances, weighted summation is much to be preferred. The ideal weights are the saturation coefficients. When these are unknown, we must take the figures already obtained by unweighted summation as pro-

[1] If figures exact to more than three decimal places are required, the more elaborate formulæ which dispense with the intercorrelations may prove to be quicker, e.g. the 'product formula,' calculated by logs, or Spearman's well-known 'summation formula,' which, however, takes more time (see *Marks of Examiners*, pp. 285–6, eqs. xxvi and xxvii, and comments). If only two or three digits are wanted, the method in the text is the quickest; and, with tables deviating widely from the hierarchical order, it yields results that are more exact.

visional estimates: they are shown under the heading of 'multipliers' in the first column of Table II.

1. Insert the self-correlations, i.e. the squares of the estimated saturation coefficients (·81, ·64, etc.), in the leading diagonal of the original table (Table I).

2. Multiply each row of correlations by the corresponding saturation coefficient (e.g. 1st row: ·90 × ·81 = ·729, ·90 × ·72 = ·648,

I. SINGLE-FACTOR ANALYSIS

A. Simple-summation Method $\left(r_{ag} = \frac{\sum_i r_{ai}}{\sqrt{\sum_j \sum_i r_{ji}}} \right)$

TABLE I

Observed Intercorrelations and Calculation of Saturation Coefficients

Sat. Coeffs.	·90	·80	·70	·60	·50	Totals.
·90	—	·72	·63	·54	·45	2·34
·80	·72	—	·56	·48	·40	2·16
·70	·63	·56	—	·42	·35	1·96
·60	·54	·48	·42	—	·30	1·74
·50	·45	·40	·35	·30	—	1·50
Totals	2·34	2·16	1·96	1·74	1·50	9·70
Self-correlations: 1st estimate	·80	·65	·50	·35	·25	2·55
1st completed totals	3·14	2·81	2·46	2·09	1·75	12·25 = 3·50²
Sat. Coeffs.* 1st estimate	·897	·803	·703	·597	·500	3·50 (checks)
Squares	·805	·645	·494	·356	·250	
Self-correlations: 2nd estimate	·81	·64	·49	·36	·25	2·55
and completed totals	3·15	2·80	2·45	2·10	1·75	12·25 = 3·50²
Sat. Coeffs.† Final estimate	·90	·80	·70	·60	·50	3·50 (checks)
Squares (for checking)	·81	·64	·49	·36	·25	2·55 (checks)

* Divisor = $\sqrt{12{\cdot}25} = 3{\cdot}50$. † Divisor = $\sqrt{12{\cdot}25} = 3{\cdot}50$.

B. Weighted Summation Method $\left(r_{ag} = \frac{\sum_i r_{ig} r_{ai}}{\sum \sum r_{ig} r_{ji} \div \sum_i r_{ig}} \right)$

TABLE II

Weighted Intercorrelations and Calculation of Saturation Coefficients

Multipliers.	Products.					Total.
·90	·729	·648	·567	·486	·405	
·80	·576	·512	·448	·384	·320	
·70	·441	·392	·343	·294	·245	
·60	·324	·288	·252	·216	·180	
·50	·225	·200	·175	·150	·125	
Totals 3·50	2·295	2·040	1·785	1·530	1·275	8·925 ÷ 3·50 = 2·55
Sat. Coeffs.*	·90	·80	·70	·60	·50	3·50
Squares .	·81	·64	·49	·36	·25	2·55 (checks)

* Divisor $= \frac{\text{Total of Products}}{\text{Total of Multipliers}} = \frac{8{\cdot}925}{3{\cdot}500} = 2{\cdot}55.$

etc.; 2nd row: ·80 × ·72 = ·576, etc.); the products are shown in the body of Table II.

3. Add each column of products to obtain its total (2·295, 2·040, etc.).

4. Add the totals of the columns to obtain their grand total (8·925).

5. Add the trial multipliers or weights (saturation coefficients) to obtain their total (3·50).

6. Divide the grand total of weighted correlations (8·925) by the total of the weights (3·50), obtaining the quotient 2·55. This should give the total of the variance contributed by the first factor.

7. Divide each weighted total by this quotient (2·295 ÷ 2·55 = ·90, etc.). The quotients so obtained give revised estimates for the saturation coefficients as obtained by weighted summation. They should be approximately equal to the trial multipliers with which we started (·90, ·80, etc.). If not, the revised saturation coefficients must be taken as trial weights, and the whole process repeated.

8. Square the final revised saturation coefficients. The sum of the squares (2·55) should give the total variance contributed by the first factor and therefore be equal to the quotient obtained above in

step 6. If it is not, each saturation coefficient must be divided by

$$\sqrt{\frac{\text{Sum of squares of revised saturation coefficients}}{\text{Total factor-variance as obtained in step 6.}}}$$

With a perfect hierarchy, as, for example, in the present case, the methods of weighted and unweighted summation give identical results. If the trial values taken for the saturation coefficients (and therefore for the self-correlations which are assumed to be the squares of the saturation coefficients) had been inaccurate, we should have taken the calculated values as giving better weights and better communalities, and started all over again. The reiterated calculations would ultimately lead to the figures given above. It will be noticed that these final figures imply values for the self-correlations which reduce the correlation table to a *perfect* hierarchy, that is, to a matrix of rank one, or, in other words, of the *lowest possible rank which the observed intercorrelations permit.* Thus, with the foregoing procedure of successive approximation, the methods of unweighted and weighted summation both yield an analysis entailing the smallest possible number of factors.

C. Bipolar Hierarchies

In earlier researches, where tests of intellectual abilities or educational attainments were applied to random samples of the population, the correlations were usually positive throughout; but when such tests are applied to more homogeneous samples (e.g. school classes instead of complete age-groups), negative correlations are often found, systematically distributed within the table. The negative values are still more conspicuous when we correlate assessments for emotional or temperamental traits instead of tests of cognition. The elimination of a general factor has much the same effect as selecting a homogeneous group; and, when the summation method has been used, the residual correlations yield a special kind of pattern which may be called 'bipolar.'

In later work on factor-analysis, therefore, modified methods proved essential in order to deal with negative correlations as well as positive ([30], [33]). In principle either simple summation or weighted summation may still be employed. The requisite modifications may best be illustrated by taking once again an artificial example.

Table III shows a perfect bipolar hierarchy, constructed from the saturation coefficients set out along the top and left-hand margins. The pattern shows certain new peculiarities. In the upper half[1] (or 'chief,' if I may borrow a convenient term from heraldry) the

[1] 'Half' as judged by the totals, not by the number of rows or columns. The 'dexter' half will be on the *left* of the spectator or reader.

coefficients in each row decrease in the same proportions, just as they do in an ordinary hierarchy that is positive throughout; but now they do not merely decrease to low positive values; they continue to decrease *algebraically*, and so become negative, with larger and larger *numerical* values. Since the whole table is symmetrical, the trend of the lower half (or 'base') will reverse that of the upper half, the 'dexter base' containing the same negatives as the 'sinister chief,' so that in the lower rows the coefficients *increase* algebraically as we move to the right or downwards.[1]

As before, the saturation coefficients are supposed to be unknown; and our object is to rediscover them. The simplest procedure is as follows:

1. Add up the totals for the positive figures and for the negative figures respectively in each column (+ 1·52, − 1·52, etc.). With tables of residuals as usually obtained, these absolute sub-totals will be nearly, if not quite, identical, and the algebraic grand totals therefore approximately zero. Hence, in its original form, simple summation can no longer be employed.

A solution to the difficulty can be found by remembering two points: (i) A 'hierarchy' (i.e. as the term is here used, a matrix of rank one) may be produced by the multiplication of a row of positive and negative saturations as well as by a row of positive saturations only. (ii) The use of simple summation is merely a

[1] As already remarked, patterns of this type do not conform with the usual definition of the Spearman hierarchy: namely, "that every coefficient is greater than any coefficient to the right of it in the same row, or below it in the same column" ([12], p. 275, [39], p. 165). When the intercolumnar criterion is applied, the correlations of columns on the left with columns on the right are negative. For this reason Stephenson and others have claimed that my bipolar patterns require two group-factors instead of a single general factor to explain them—a perfectly permissible hypothesis ([96], p. 349). On the other hand, the bipolar pattern *does* obey the proportionality criterion, and therefore the usual test for a matrix of rank one—Spearman's 'tetrad-difference criterion'—provided we are allowed to take the 'tetrads' as including negative figures. Spearman himself, however, considers negative saturations meaningless.

It will also be noted that, if we ignore the pattern of signs, the numerical values can be analysed as if they formed an ordinary hierarchy of the exclusively positive type. Then, on prefixing appropriate signs to the saturation coefficients so obtained, we reach the same results as below. Where we are dealing with a single factor alone, this procedure is justifiable, for we may legitimately reverse the order of measurement of the testees for those tests which would otherwise have negative saturations. When we are dealing with a table of residuals, and still more with a superimposed succession of such tables, this device seems to me to involve several difficulties, both theoretical and practical. Nevertheless, even factorists who ultimately admit negative saturations into their factorial matrices recommend the student to convert negative correlations into positive so far as possible, and thus keep all the saturation coefficients positive until the final stage (cf. Thurstone, *loc. cit.*, p. 112; Guilford, *loc. cit.*, p. 486). In practice this leads to considerable perplexity when the beginner has to analyse a succession of residual tables and insert the signs as after-thoughts. If we recognize that the method of simple summation (or centroid method, as Thurstone would call it) is merely a simplified substitute for that of weighted summation, the mode in which the signs require to be altered becomes obvious: for the reversals now appear as an essential part of the weighting process, not a series of arbitrary dodges to assist the computer.

simplified substitute for weighted summation. In the latter case, the weights are the saturation coefficients, and with a bipolar hierarchy a number of them will consequently be negative: on multiplication, therefore, the minus sign of a negative weight will evidently reverse the signs of the residuals throughout the row, thus converting their minus signs to plus, and *vice versa*. With simple summation, we tacitly assume that the numerical value of each weight is unity throughout; but *we must retain the negative signs* of these weights wherever such signs arise. The only problem, therefore, is to determine in advance which rows will have negative weights. Once again the best guide is the pattern of the correlation table.

2. Accordingly, begin by rearranging the residuals, so that the whole table approximates as closely as possible to a bipolar hierarchy. This means that as many negative signs as possible must be collected into a single oblong block, which may be conveniently kept in the north-east quarter[1] of the table: owing to symmetry, the same block will automatically appear (duly transposed) in the south-west quarter.[1] By maximizing the number of negative signs (or rather, what is more important, the total amount of negative correlation in each column) within the north-east and the south-west blocks, we simultaneously maximize the number of positive signs (or rather the total amount of positive correlation) within the north-west and the south-east blocks.

Table III shows the resulting pattern. With an empirical table of figures, slight irregularities may be found which render the rearrangement a little more complicated. In such cases the following hints may be found useful:

(*a*) Start by moving the tests with the largest number of positive correlations towards the beginning (i.e. left) of the table, and the tests with the largest number of negative correlations towards the end (i.e. right).

(*b*) Pick out the biggest negative correlation (— ·72); the tests thus negatively correlated should, as a rule, appear at the top and at the bottom of the hierarchy respectively.

(*c*) Next pick out the two tests having the largest positive correlation with either of these (·48 and ·63); they will usually have large negative correlations with the other of these (— ·54, — ·56), and with each other (— ·42), and therefore appear next to the top and next to the bottom test in the hierarchy respectively.

(*d*) Continue in this way, working from the three extreme corners of the table towards the centre.

[1] 'Quarter' as judged by the totals: cf. footnote 1, p. 454 above.

C. Bipolar Hierarchy

TABLE III

Rearranged Correlations and Calculation of Saturation Coefficients

Weights.	Sat. Coeffs.	+·80	+·60	+·40	+·10	−·30	−·70	−·90	Totals.
+1	+·80	+[·64]	+·48	+·32	+·08	−·24	−·56	−·72	·00
+1	+·60	+·48	+[·36]	+·24	+·06	−·18	−·42	−·54	·00
+1	+·40	+·32	+·24	+[·16]	+·04	−·12	−·28	−·36	·00
+1	+·10	+·08	+·06	+·04	[·01]	−·03	−·07	−·09	·00
−1	−·30	+[−]·24	+[−]·18	+[−]·12	+[−]·03	−[+][·09]	−[+]·21	−[+]·27	·00
−1	−·70	+[−]·56	+[−]·42	+[−]·28	+[−]·07	−[+]·21	−[+][·49]	−[+]·63	·00
−1	−·90	+[−]·72	+[−]·54	+[−]·36	+[−]·09	−[+]·27	−[+]·63	−[+][·81]	·00
Totals:									
Upper half .		+1·52	+1·14	+·76	+·19	−·57	−1·33	−1·71	0·00
Lower half .		−1·52	−1·14	−·76	−·19	+·57	+1·33	+1·71	0·00
After weighting		+3·04	+2·28	+1·52	+·38	−1·14	−2·66	−3·42	[+]14·44 = 3·80²
Sat. Coeffs.* .		+·80	+·60	+·40	+·10	−·30	−·70	−·90	[+] 3·80

* Divisor = $\sqrt{14{\cdot}44} = 3{\cdot}80$.

N.B.—The signs in brackets are the original *signs: those in front of them are the effect of multiplication by the 'weights.'*

(*e*) The numerical totals of the columns will help to decide the final order: tests having the largest totals and sub-totals should be placed towards the beginning or towards the end of the table, and generally the absolute totals should first descend and then rise again: (2 × 1·52, 2 × 1·14, . . ., 2 × 0·10, 2 × 0·57, 2 × 1·71).

3. Now insert estimates for the self-correlations[1] in the leading diagonal: (·64, ·36, etc.). Usually the figures at either end of the diagonal (·64, ·81) should be larger than the largest correlation in the column, and those near the centre (·01, ·09) nearly as small as, if not smaller than, the smallest correlations. Being squares, or the sums of squares, all the self-correlations will be positive. These trial estimates will usually have to be checked and revised as explained in the previous example.

4. Prefix appropriate weights, + 1 or − 1, to each of the rows in turn, + 1 in front of the upper rows (containing the north-east block of negative correlations), − 1 in front of the lower rows (containing the same block in the south-west corner).[2] If there is occasionally some doubt as to the appropriate sign (e.g. in the middle rows), the criterion is simple: the sign of the weight for the *x*th row should be the same as the sign of its weighted sum (i.e. of the weighted sum for the *x*th column).

5. Multiply each row by the sign thus prefixed: i.e. reverse all the signs in those rows which have negative weights. The effect is shown in Table III by bracketing the original signs (nearest the figure) and prefixing the new sign: (− ·24 becomes + ·24; + ·09, although a self-correlation, becomes − ·09 etc.).

6. Using the new signs, add each column[3] to find its algebraic total: (+ 3·04, + 2·28, . . ., − 3·42). The sign of the total for each column should agree with the sign employed to weight the corresponding row: if not, the sign of the weight must be changed, and the multiplication and addition (steps 5 and 6) repeated. The method thus provides its own check.

7. Find the numerical or absolute total of these column-totals: i.e. treat them all as positive, and add them up: (3·04 + 2·28 + . . . + 3·42 = 14·44).

8. Find the square root of this grand total: ($\sqrt{14{\cdot}44} = 3{\cdot}80$).

9. Divide the totals of the columns, with their original signs, positive or negative as the case may be, by this square root:

[1] With tables of residuals there will already be a figure in each diagonal cell, possibly a negative figure: after using them for the check described in step 1, they should, particularly if merely based on previous *estimates*, be discarded for better estimates.

[2] It is algebraically indifferent whether we assign the negative values to the lower or to the upper rows, so long as the signs are opposite: hence it may sometimes be more appropriate to determine the choice by the psychological nature of the bipolar factor.

[3] Here (but not always) it will be sufficient merely to add the positive and negative sub-totals.

(+ 3·04 ÷ 3·80 = + ·80 ; . . . ; − 3·42 ÷ 3·80 = − ·90). The quotients should be the saturation coefficients required.

With this and other bipolar tables it is instructive to continue with the full method of weighted summation, using the saturation coefficients as weights, and repeating the process by successive approximation as described above (section B, pp. 451–3). In this case, as in the preceding example, it will be found that the saturation coefficients remain unchanged.

II. MULTIPLE-FACTOR ANALYSIS

Where a given table of correlations appears to be produced by more than one factor, we have to split the matrix of observed coefficients into the sum of two or more simple 'hierarchies' (matrices of rank one). Each of these component hierarchies may cover either (1) the entire series of tests (in which case the factor is described as a 'general factor'), or (2) a limited group of tests only (in which case it is called a 'group-factor').

(1) *GENERAL-FACTOR METHODS*

The former procedure consists essentially in applying (i) the ordinary single-factor analysis to the correlations actually observed in order to determine the first factor, and then (ii) the modified bipolar analysis to the successive tables of residual correlations in order to determine the remaining factors.

The hypothesis of summed hierarchies now implies that the variance for each test (its 'self-correlation,' as it is sometimes inaccurately termed) should be the sum of the *squares* of its factor-saturations, just as the intercorrelation between any two tests is the sum of the *products* of their factor-saturations.

With a table of covariances the calculations are perfectly straightforward. If n is the number of tests and r the number of factors, we have (in general) $\frac{1}{2}n\,(n+1)$ independent values in the covariance matrix, and $nr - \frac{1}{2}r\,(r-1)$ independent parameters to calculate in the factorial matrix.[1] Hence, if the variances are known, there will be $r = n$ factors, of which, of course, only the

[1] As will be seen from the procedure described below, there are n degrees of freedom for the first factor; only $(n-1)$ for the second (since the totals must be zero); only $(n-2)$ for the third; and only $(n-p+1)$ for the pth factor, since in general the totals must be zero, not only as they stand (i.e. when weighted by the signs of the first factor which are all positive), but also when weighted by the signs of *any* of the $(p-1)$ preceding factors. This latter point can easily be proved algebraically (see *Notes*), and may be verified from the tables of saturations given below: it will also be found to hold good approximately for the tables analysed by Thurstone's centroid method (e.g. the five columns of saturations in his table 25, *Vectors of the Mind*, p. 117; the slight discrepancies are due to the substitution of fresh values for the diagonal residuals in the case of the later factors).

first few will usually be statistically significant. With a table of intercorrelations, where the variances or 'self-correlations' are unknown, we have only $\frac{1}{2}n(n-1)$ independent values in the correlation matrix. We have, therefore either (i) to make some convenient but arbitrary assumption about the nature of these variances, e.g. that they are all equal and standardized at unity: in that case we should obtain n factors—most of which could have no real statistical significance and would be due solely to the additions made to the true variances to bring them all up to unity. Or, alternatively, we can (ii) follow the procedure described above, and insert an estimated 'self-correlation' to fit the requirements of each successive hierarchy solely. But now, if there are more than one factor, the total variance (i.e. the self-correlation used in calculating the first factor) should be at least equal to the sum of the self-correlations for *all* the factors, and generally the variances required for calculating for the pth factor should each be the sum of contributions of the $(r-p+1)$ factors that remain to be calculated. This involves readjustments in the estimated self-correlations, reduces the number of ascertainable factors, and leads to a lengthy process of successive approximation. It yields, in the end, not n factors, but the minimum number required to account for the intercorrelations as given, or, in technical language, a factorial matrix of minimum rank. Actually, however, the figures given for the intercorrelations are themselves approximate abridgments for irrational fractions. Hence, with large correlation tables, there is little point in seeking a factorial matrix having *exactly* the minimum rank deducible from the figures given; and it is even arguable that some small allowance should be included for variance due to the 'specific factor' or rather to that part of the test that does not overlap with any other test in the table.

In these circumstances the working procedure I have suggested assumes that the variances for the different processes tested will differ in the main according to their complexity, and that (in default of other evidence) this can be estimated from the maximum covariance for which each could be responsible (see p. 286). The inserted figures will therefore be (i) not less than the total variances that would be obtained if the successive approximations were carried through to the end, but (ii) never so large as to make the total variances exactly or even approximately equal.[1]

By way of illustration let us now take actual figures; I choose a set of correlations printed and analysed in my 1917 *Report*.[2] For simplicity I shall here confine the analysis to six tests only, namely, Composition, Reading, Spelling, Handwork, Writing, and Drawing, and seek figures correct to two decimal places. A more detailed analysis by the summation method, and a full discussion of the inferences to be drawn, may be found in the *Report* itself.

[1] The insertion of the reliability coefficients usually makes the proportionate variances approximately equal, and tends unduly to diminish the variances of the more complex tests, since the simplest tests usually have the highest reliability.

[2] *Loc. cit.*, Table XVIII, p. 52. I am indebted to Miss G. Bruce for checking the calculations: a few of the figures originally printed have required correction, particularly those for Writing, where two tests have here been amalgamated into one.

Table IV reproduces the intercorrelations actually observed. With only 6 tests, the correlation table should be completely explicable in terms of

$$r = \tfrac{1}{2}\{(2n + 1) - \sqrt{8n + 1}\} = 3 \text{ factors.}^1$$

A. Simple Summation Method

I shall repeat the working instructions in full, partly for convenience of reference and partly because of certain special difficulties that an actual problem inevitably entails.

(i) *First Factor*

Formula : Saturation coefficient $r_{ag} = \dfrac{\Sigma_i r_{ai}}{\sqrt{\Sigma_i \Sigma_j r_{ji}}}$ as before.

1. Find the total intercorrelation of each test (2·19, 1·48, etc.), by adding each column of observed coefficients.

2. With the aid of these totals rearrange the table so as to exhibit the general trend or pattern as clearly as possible. The guiding principles are, first, to bring correlations of the same size as near together as possible, and secondly to keep the changes of size along the rows or columns moving in the same direction and in parallel directions, as smoothly and as continuously as possible. Here, if the student begins by attempting to arrange the figures given above in plain hierarchical order, he will at once discover that this simpler pattern is disturbed by the low correlations between the verbal group of tests and the manual group. Consequently, although all the correlations are positive, we have to recognize at least two high points of correlation instead of one, namely, at the bottom right-hand corner as well as at the top left. This in turn indicates that the totals should be arranged so as first to descend and then to rise again.

We take the totals for (1) Composition, Reading, and Spelling, in descending order, and those for (2) Handwork, Writing, and Drawing, in ascending order. The rearrangement brings out an effect that I have described as ' cyclic overlap ' : [2] the correlations near the leading diagonal tend to be higher than simple hierarchical arrangement would require, and they also tend to rise, not only

[1] This formula follows from the considerations set out on the preceding page: for proof, see p. 109.

[2] *Loc. cit.*, p. 59. If the student finds some difficulty in making a satisfactory rearrangement, he may be content with a first approximation to begin with. Later we shall see from the *completed* totals (last line but one) that it would have been better to place Drawing at the end instead of Writing; and this is confirmed by the totals of the residuals (Table IV).

II. MULTIPLE-FACTOR ANALYSIS

A. Simple Summation

TABLE IV

Observed Intercorrelations and Calculation of Saturations for First Factor

Tests.	Comp.	Read.	Spell.	Hand.	Draw.	Writ.	Totals.
Comp. . .	—	·58	·49	·30	·38	·44	2·19
Read. . .	·58	—	·46	·10	·13	·21	1·48
Spell. . .	·49	·46	—	·09	·12	·25	1·41
Hand. . .	·30	·10	·09	—	·50	·28	1·27
Draw. . .	·38	·13	·12	·50	—	·36	1·49
Writ. . .	·44	·21	·25	·28	·36	—	1·54
Totals . .	2·19	1·48	1·41	1·27	1·49	1·54	9·38
Estimated Communalities .	·70	·60	·40	·40	·60	·40	3·10
Completed totals	2·89	2·08	1·81	1·67	2·09	1·94	12·48
Sat. Coeffs.* .	·818	·589	·512	·473	·592	·549	3·533

* Divisor $= \sqrt{12{\cdot}48} = 3{\cdot}533$.

towards the top left- and bottom right-hand corners, but also towards the top right and the bottom left.

3. Estimate in round figures the diagonal self-correlations by smoothing off the pattern as before. Owing to the double grouping and the cyclic overlapping just described, the more complex pattern produced by superposed factors usually suggests a diagonal ridge, high towards the corners, especially towards the top left, and sinking in the centre towards the diagonal trough that crosses it. Accordingly, for the space in the 1st column take the highest correlation in that column (·58) and increase it appreciably (say to ·70); for the space in the 2nd column, increase the highest figure (·58) only slightly (say to ·60); for the spaces in the 3rd and 4th, take the highest adjacent figure, and diminish it; for the spaces in the 5th and 6th, take the highest of the adjacent figures, and increase them.[1]

[1] Note once again that inserting the highest correlations at every point would obviously be wrong. It would tend to underestimate the variance for a test like Composition and would overestimate very considerably the variance of tests like Spelling or Handwork. For Handwork and Drawing, the same high figure, ·50, would be taken as the variance for both; but it is evident that the contribution of Handwork to the variance and covariance of the whole table must be much smaller than that of Drawing.

4. Add each estimated figure to the total of the corresponding column: (2·19 + ·70 = 2·89, etc.).

5. Add these completed totals: (2·89 + 2·08 + . . . = 12·48).

6. Find the square root of this grand total: ($\sqrt{12{\cdot}48} = 3{\cdot}533$).

7. Divide the augmented total of each column by this square root: (2·89 ÷ 3·533 = ·818, etc.).

The quotients (last line of Table IV) should yield approximate values for the saturation coefficients for the first factor. It is no longer possible to check them straight away by comparing their squares with the estimated self-correlations, since the latter are intended to include the contributions, not of the first factor only, but of all the factors; but the differences should seldom be large.

(ii) *Second Factor*

Formulæ: Residual correlation, $r'_{ab} = r_{ab} - r_{ag}\, r_{bg}$

$$\text{Saturation coefficient, } r'_{as} = \frac{\sum_i r'_{ai}}{\sqrt{\sum_j \sum_i r'_{ji}}}$$

8. Multiply each of the saturation coefficients by each of the others, as before, thus obtaining a perfect hierarchy of theoretical correlations, attributable to the first factor, as shown in Table V: (·818 × ·818 = ·669, ·818 × ·589 = ·482, etc.).

9. Enter the observed figures above them, and subtract the theoretical figure from the observed, as shown: (·70 − ·669 = ·031, etc.). The remainders will form a table of residuals, with both positive and negative signs. With the general-factor method, we have now to factorize this entire table of residuals as a single matrix.[1] If we add the residuals as they stand, we shall find that the total of each column, like the total of the deviations about an average (or rather an average gradient), comes exactly to zero. We must therefore adopt the procedure described above for factorizing a *bipolar* table: (cf. Table III, there is no need to print a fresh table to illustrate the working).

10. We begin as usual by rearranging the residual intercorrelations so as to bring out the general pattern. We may assume that this is approximately hierarchical, and quartered into positive and

[1] With an actual table the residuals should first be tested for statistical significance. Here N = 120; and 4 out of the 15 residuals are significant. χ^2 (see p. 339) = 37·5; hence P < 0·0001 ([110], p. 540). The s.d. of the ratios (residual ÷ s.e. of correlation) is 1·60—well over the theoretical value of 1·00 (this test gives a slightly higher value for P). As we have seen (p. 368) with simple summation, the use of χ^2 is not strictly valid: but here the least-squares procedure leads to virtually the same results (p. 474).

TABLE V

Hierarchy for First Factor and Calculation of Residuals

Tests.	Comp.	Read.	Spell.	Hand.	Draw.	Writ.	Total.
Sat. Coeff. .	·818	·589	·512	·473	·592	·549	—
Obs. Corr. .	·70	·58	·49	·30	·38	·44	—
·818 . . .	·669	·482	·419	·387	·484	·449	—
Residuals . .	·031	·098	·071	—·087	—·104	—·009	·000
Obs. Corr. .	—	·60	·46	·10	·13	·21	—
·589 . . .	—	·347	·301	·278	·349	·323	—
Residuals . .	·098	·253	·159	—·178	—·219	—·113	·000
Obs. Corr. .	—	—	·40	·09	·12	·25	—
·512 . . .	—	—	·263	·243	·303	·281	—
Residuals . .	·071	·159	·137	—·153	—·183	—031	·000
Obs. Corr. .	—	—	—	·40	·50	·28	—
·473 . .	—	—	—	·223	·280	·259	—
Residuals . .	—·087	—·178	—·153	·177	·220	·021	·000
Obs. Corr. .	—	—	—	—	·60	·36	—
·592 . . .	—	—	—	—	·349	·325	—
Residuals . .	—·104	—·219	—·183	·220	·251	·035	·~~100~~
Obs. Corr. .	—	—	—	—	—	·40	—
·549 . . .	—	—	—	—	—	·303	—
Residuals . .	—·009	—·113	—·031	·021	·035	·097	·000
Total residuals .							
Upper half .	·200	·510	·367	—·418	—·506	—·153	·000
Lower half .	—·200	—·510	—·367	·418	·506	·153	·000
Abs. Sum . .	·400	1·020	·734	·836	1·012	·306	—

negative rectangles arranged symmetrically about the leading diagonal, as described on pp. 456–8, step 2.[1] Follow the suggestions therefore for marshalling the coefficients in this order. If the table of observed correlations was successfully rearranged at the very outset (step 2, p. 461), the amount of further rearrangement will probably not be very great. The essential result will be two (possibly more) square blocks of positive coefficients astride the main diagonal, and rectangular blocks of negative correlations in the N.-E. and S.-W. quarters (as shown by the residuals in Table V).

11. For the arithmetical check described in step 9, we retained the calculated residuals in the leading diagonal. But these were the result of the first rough estimates in step 3: they may even be negative. Discard them; and substitute estimates for the self-correlations or residual 'variances,' positive throughout, to fit the general pattern.

12. Assign to each row its appropriate weight, + 1 or − 1. For this purpose first decide on the block of (mainly) negative residuals, attributable to the multiplication of negative and positive saturations. Owing to the symmetry of the table, this negative block will appear twice. Its longer horizontal rows (in the lower quarter) will usually receive a weight of − 1; its shorter horizontal rows (in the upper quarter) a weight of + 1 (cf. p. 457).

13. Multiply each row by the weighting-sign thus prefixed, i.e. where the weight is negative, reverse the signs throughout the whole row from left to right of the table.

14. Using the new signs, add each column to find its algebraic total. See that the sign of each total confirms the sign already employed to weight the corresponding row.

15. Find the numerical or absolute grand total of these column-totals: i.e. treat them as all positive, and add them up.

16. Find the square root of this grand total.

17. Divide the total of each column (with its original sign, positive or negative as the case may be) by the square root. The quotients should yield the saturation coefficients for the second factor. As a check on the working, note that their total should be approximately zero: (exactly zero, as in Table XI below, had we not modified the diagonal residuals).

(iii) *Remaining Factors*

18. To obtain saturation coefficients for the third and other factors, if necessary, we continue exactly as before, calculating the

[1] If there are more than one residual factor, these 'grand quarters' will be 'counter-quartered'; hence only an approximate hierarchy will be attainable at this stage.

second-factor hierarchy, subtracting it from the residuals left by the first factor (with their original signs), testing the significance of the new residuals, rearranging them in bipolar order so as to obtain an approximate hierarchy of positive and negative quarters, estimating figures for the diagonal, adding the columns, and dividing by the square root of the grand total—as just described. In theory this procedure should be continued with one set of residuals after another, until the final residuals are virtually zero for the number of decimal places to which we are working: in practice, it is unnecessary to calculate detailed figures for non-significant residuals, i.e. as a rule, for more than three or four factors.

19. As a final check, square *all* the saturation coefficients and find the sums for each test, if necessary adding in a small estimated fraction for any test that still shows an appreciable residual which has not been explicitly factorized. The sum of the squares for each test should be approximately identical with its variance ('communality' or 'self-correlation') as estimated at the very outset.

With the present set of correlations the results obtained by the foregoing procedure are shown in Table VII, A, i. They fit the intercorrelations reasonably well to two decimal places. Strictly, however, they must be regarded as first approximations only. To obtain more accurate figures, we should take the sum of the squares of the saturations for each test (·710, ·614, . . ., ·389) and insert it (or a figure still further increased or reduced) in place of the 'estimated self-correlation' (·70, ·60, . . ., ·40) originally used for obtaining the 'completed totals' in Table IV; with these new estimates we should then have to repeat the whole process.

After two or three such repetitions I reach the following figures for the variances or self-correlations:

·7216, ·6351, ·3794, ·3994, ·6266, ·3924

The saturations ultimately obtained are given below in Table VII, A, ii. The residuals remaining after extracting these three factors are all less than ·001.[1] It is clear that, if we worked to further decimal places, we could *approximate as closely as we wished to the given intercorrelations with three factors only.* In practice, of course, accuracy to four decimal places is seldom required: here,

[1] This, of course, could be predicted at the outset (p. 461). With 6 tests we have 15 intercorrelations. If the 6 self-correlations or variances are themselves to be determined from these 15 intercorrelations, then there are only 15 degrees of freedom for determining the saturations. To determine the 6 saturations for the first factor, 6 degrees of freedom will be required: to determine the 6 saturations for the second factor, only 5 (having determined 5 saturations the 6th follows automatically, because the total is by hypothesis zero); to determine the 6 saturations for the third factor, only 4 will be required. But $6 + 5 + 4 = 15$. Hence no more factors can be determined unless the variances include some arbitrary quantities.

as can be seen from Table VII, A, i and ii, the improvements in the saturation coefficients produced by successive approximation have nearly all been less than ·01.[1]

B. Weighted Summation

If we fill in the leading diagonal with the figures for the 'communalities' derived by the method of simple summation, we can go on (if we wish) to improve the factorial matrix by the method of weighted summation. Where the diagonal values are independently determined and already known (as in a table of covariances), the method of weighted summation should always be used. But where (as in most large correlation tables) the diagonal values can themselves consist only of estimates subject to appreciable correction, the further improvement thus secured should be regarded rather as a check than as a final or indispensable aim. If, however, additional calculations are to be based on the results, or the χ^2 test of significance is to be applied, then the improvement is not only desirable but essential. The figures eventually reached by weighted summation are equivalent to what in matrix algebra would be termed the 'latent roots' and the 'latent vectors' of the matrix; and possess all the useful properties of such quantities.[2]

To illustrate the calculations required by weighted summation let us take the same correlation table.

In its application to a perfect hierarchy, as we have already seen (p. 453), the principles involved are obvious enough. But when the table departs from the hierarchical arrangement, they are hardly self-evident. Their theoretical basis has been indicated in the text above (p. 320). The argument is briefly as follows. (*a*) Any matrix, whatever its rank (i.e. however much it may depart from a perfect hierarchy) can nevertheless be reduced to as perfect a hierarchy as may be desired by repeated self-multiplication ('table-

[1] The results obtained by taking the 'communalities' to be equal to the largest correlation in the column are far less accurate. For example, for the first factor we obtain for Composition a saturation of ·784 and for Handwriting one of ·501: on comparing these with the 'true' values (as obtained by prolonged approximation and the method of least squares, Table VII, B) it will be seen that they diverge far more than the saturations obtained with the first rough attempt to estimate the 'communalities' from the pattern.

[2] In quantum mechanics these quantities are usually known as the *Eigenwerten* (or 'characteristic values,' which correspond to our 'factor-variances') and the *Eigenvektoren* (or 'characteristic vectors,' which correspond to our normalized 'factor saturations'): cf. Burt, [101], p. 77 f., [115]. It may be noted that, if instead of the 'communalities' we insert unity throughout the leading diagonal, the results obtained by weighted summation should be identical with those reached by Hotelling's 'method of principal components' and by Kelley's 'rotation method,' though Kelley would prefer to insert 'reliabilities': (cf. [102] for illustrative comparisons).

by-table' multiplication) (Burt, [102]): the saturation coefficients for the first factor (or their relative proportions) can then be determined by simply adding the columns as before. (*b*) But instead of summing columns *after* the multiplications, we may sum them at the start; we shall then have to multiply by a single row (or column) only, instead of by the whole table ('table-by-column' multiplication). (*c*) After the first 2 or 3 multiplications each further multiplication of the matrix by itself yields sums that are equivalent to the figures that would be obtained by multiplying the product sums by the first factor-variance (i.e. by the largest). At each stage, therefore, we must divide the results by an estimate for that variance. And once again we may either divide the weights by the variance *before* multiplying or divide the products by the variance *after* multiplying.

These considerations lead to several alternative working procedures, each of which has its own special merits as regards speed or exactitude. 'Table-by-table' multiplication, for example, will entirely avoid the need for successive approximation. It is, however, too arduous a method for general use; and will not, therefore, be included here.[1]

(i) *Table-by-Column Multiplication: Full Procedure (with Geometrical Progressions).*—For highly accurate work the sums of the columns should be taken as the first weights just as they stand. The repeated table-by-column multiplication then yields figures whose diminishing differences are approximately in geometrical progression. The successive differences can be summed to infinity, and an exceedingly close approximation obtained, running to 6 or 8 decimal places, with comparatively little labour.

Except for purely theoretical investigations, so high a degree of accuracy is hardly ever required: the procedure itself has already been sufficiently illustrated in chapter xiii, Table VI, p. 326.

(ii) *Abridged Method (with Ratios of Variance as Weights).*—Where a more rapid procedure and less extended figures are required, many of the intermediate multiplications can be skipped, the computer guessing the trend of the progressions without full or formal calculation. The multiplications are repeated until two consecutive sets of product-sums agree in their proportions with sufficient closeness.

To compare any set of product-sums with the preceding set, the new set must be divided, as we have seen, by the variance. Alternatively we can estimate the variance by dividing the total of the product-sums by the total of the saturation coefficients which we

[1] A more detailed account, with examples, is given in my paper on 'Methods of Factor-analysis with and without Successive Approximation' [102].

have used as weights, or (with less accuracy) by dividing any single product-sum by the corresponding single saturation coefficient. If we carry out the latter division *before* multiplying, the weight when divided by the corresponding saturation coefficient becomes equal to unity, since the saturation coefficient is the weight.[1]

Let us begin by following this procedure, since it renders the essential comparison more obvious to the eye.

(*a*) *1st Approximation* (*Table VI*, *a*)

1. We begin, as in simple summation, by inserting the estimated values for the self-correlations (·7216, ·6351, etc.),[2] and adding up the several columns as they stand, i.e. giving, for the first approximation, an *equal* weight—namely, unity—to each row.

2. Instead of converting the totals straight away into saturation coefficients, each total is simply divided by the largest in the set: (2·91 ÷ 2·91 = 1·00; 2·12 ÷ 2·91 = ·73; etc.): though we can no longer make *all* the weights equal to unity, we can at least make the *largest* equal to unity, thus simplifying both comparisons and calculations. These ratios are now to be used as provisional weights for the next approximation. We cannot presume that they are correct to more than the first decimal place. Hence to save labour we merely take 1·0, ·7, ·6, etc. We then proceed as in Table VI, *b*.

(*b*) *2nd Approximation* (*Table VI*, *b*)

3. (*a*) Multiply each row of correlations by the corresponding new weight: (·7216 × 1·0 = ·722, etc.; ·58 × ·7 = ·406; etc.). (*b*) Add each column of products: (·722 + ·406 + . . . + ·308 = 2·176; etc.). The largest of the totals (2·176) will form a first

[1] The theory underlying these numerical methods is succinctly set out in Frazer, Duncan, and Collar, *Elementary Matrices*, 1938, pp. 133–144. Historically, most of the iterative methods of solving the 'characteristic equation' of matrices appear to be developments of Bernouilli's method of solving polynomial equations (equivalent to repeated premultiplication of the related matrix by an arbitrary row-vector): in practice they have been employed more particularly to deal with problems arising from the principle of least squares. To psychologists the best-known use of the method is Hotelling's iterative procedure for determining 'principal components': as already noted, however, my own application differs somewhat from Hotelling's and also from the earlier iterative method of Thurstone (whose method of 'principal axes' apparently "served as a point of departure for (Hotelling's) paper" [79] p. 48; [84], p. 120). Other procedures for obtaining latent roots and vectors will be found in mathematical books and papers on the subject (for refs. cf. [81]). Usually, however, they are devised to deal with small matrices to a high degree of accuracy; and, being in addition of more general application, prove too elaborate for the needs of the computing psychologist. The advanced mathematician will find a technical summary of methods in a paper by Aitken [111], in which he has introduced many ingenious devices of his own.

[2] The figures adopted are the sums of the squares of the saturations ascertained by simple summation with successive approximation (Table VII, A, ii): since they yield a matrix of minimal rank, and since the method of weighted summation will yield a matrix of the minimal rank, the same variances hold good for both methods (p. 449)

approximation to the factor-variance: all the totals are to be expressed as ratios of this value. Accordingly—

4. Divide each total once again by the largest in the set: (2·176 ÷ 2·176 = 1·000; 1·595 ÷ 2·176 = ·735; etc.). Instead of using these figures for weights as they stand, we guess ahead and

B. Weighted Summation

TABLE VI

Calculation of Saturation Coefficients for First Factor by Successive Approximation

(Ratios of Variance as Weights)

(a) 1st Approximation

Multipliers.	Observed Correlations.					
1·0	·7216	·58	·49	·30	·38	·44
1·0	·58	·6351	·46	·10	·13	·21
1·0	·49	·46	·3794	·09	·12	·25
1·0	·30	·10	·09	·3994	·50	·28
1·0	·38	·13	·12	·50	·6266	·36
1·0	·44	·21	·25	·28	·36	·3924
Divide by 2·91)	2·91	2·12	1·79	1·67	2·12	1·93
Ratios . .	1·00	·73	·62	·57	·73	·66
Say . .	1·0	·7	·6	·6	·7	·7

(b) 2nd Approximation

Multipliers.	Products.					
1·0	·722	·580	·490	·300	·380	·440
·7	·406	·445	·322	·070	·091	·147
·6	·294	·276	·228	·054	·072	·150
·6	·180	·060	·054	·240	·300	·168
·7	·266	·091	·084	·350	·439	·252
·7	·308	·147	·175	·196	·252	·275
Divide by 2·176)	2·176	1·599	1·353	1·210	1·534	1·432
Ratios . .	1·000	·735	·622	·556	·705	·658
Adjusted .	1·00	·75	·63	·54	·68	·65

(*c*) *3rd Approximation*

Multipliers.	Products.						Total of Squares.
1·00	·7216	·5800	·4900	·3000	·3800	·4400	
·75	·4350	·4763	·3450	·0750	·0975	·1575	
·63	·3087	·2898	·2390	·0567	·0756	·1575	
·54	·1620	·0540	·0486	·2157	·2700	·1512	
·68	·2584	·0884	·0816	·3400	·4261	·2448	
·65	·2860	·1365	·1625	·1820	·2340	·2551	
Divide by 2·1717)	2·1717	1·6250	1·3667	1·1694	1·4832	1·4061	
	1·0000	·7483	·6293	·5385	·6830	·6475	
Squares . .	1·0000	·5599	·3960	·2900	·4665	·4193	$3·1317 = 1·7697^2$
Multiply by $\sqrt{2·1717}$.	·5651	·4229	·3556	·3043	·3859	·3659	
Sat. Coeffs. .	·8328	·6232	·5240	·4484	·5687	·5392	
Squares . .	·6935	·3884	·2746	·2011	·3234	·2907	$2·1717 = 1·4737^2$

(*d*) *Verification* (*Saturation Coefficients as Weights*).

Multipliers.	Products.						Total.
·833	·6011	·4831	·4082	·2499	·3165	·3665	
·623	·3613	·3957	·2866	·0623	·0810	·1308	
·524	·2567	·2410	·1988	·0472	·0629	·1310	
·448	·1344	·0448	·0403	·1789	·2240	·1254	
·569	·2162	·0740	·0683	·2845	·3565	·2048	
·539	·2371	·1132	·1347	·1509	·1940	·2115	
Totals 3·536	1·8068	1·3518	1·1369	·9737	1·2349	1·1700	7·6741
Sat. Coeffs.*	·8325	·6228	·5238	·4486	·5690	·5391	
Squares .	·6930	·3879	·2743	·2012	·3237	·2906	2·1707

* Divisor $= \frac{7·6741}{3·536} = 2·1703.$

roughly estimate the changes *they* in turn will produce. Thus, with the second column, the first computation gave ·73, the second ·735; the increase will give a large weight to a large figure; hence we guess that the third or fourth computation will give ·75. In the third column this increased weighting will probably raise ·621

to nearer ·63 in the second column. In the fourth column, the first computation gave ·58, the second reduced it to ·556, and we guess that the third or fourth will reduce it still further to (say) ·54. And so for the remaining columns.

(c) 3rd Approximation (Table VI, c)

5. (*a*) Multiply each row of the original correlations once again by these readjusted weights: (·7216 × 1·00 = ·7216, etc., as before; ·58 × ·75 = ·4350; etc., etc.). (*b*) Add each column of products as before. The largest of the totals (2·1717) yields a slightly closer approximation to the factor-variance; and once again all the totals are to be expressed as ratios of this value. Accordingly—

6. Divide each of the new totals as before by the largest in the set: (2·1717 ÷ 2·1717 = 1·000; 1·6250 ÷ 2·1717 = ·7483; etc.). The new ratios agree with the preceding to two decimal places; and, to shorten the illustration, we may suppose that they are sufficiently exact for our purpose.

We now convert the ratios into saturation coefficients. The requisite calculation is obvious when we remember that the sums of the squares of the saturations should yield the variance, which we have already ascertained. The beginner may conveniently make the calculation in two steps.

7. First normalize the ratios, so that their squares add up to unity: i.e. (*a*) square each ratio: ($1{\cdot}0000^2 = 1{\cdot}0000$; $\cdot7483^2 = \cdot5599$; etc.); (*b*) add the squares: (1·0000 + ·5599 + . . . + ·4193 = 3·1317). (*c*) find the square root of this total ($\sqrt{3{\cdot}1317} = 1{\cdot}7697$); (*d*) divide each unsquared total by this square root: (1·0000 ÷ 1·7697 = ·5651; ·5599 ÷ 1·7697 = ·4229; etc.); (*e*) check the results by squaring and adding the squares, to ensure that their total is unity: ($\cdot5651^2 + \cdot4229^2 + \ldots + \cdot3659^2 = 1{\cdot}0000$).

8. (*a*) Find the square root of the variance: ($\sqrt{2{\cdot}1717} = 1{\cdot}4737$). (*b*) Multiply each of the quotients obtained in 7 (*d*) by this square root: (·5651 × 1·4737 = ·8328; etc.). The practised calculator will condense the last two sets of computations into a single step, and simply multiply by $\sqrt{\frac{2{\cdot}1717}{3{\cdot}1317}} = \sqrt{\cdot69348} = \cdot83276$.

The products finally obtained (·8328, ·6232, etc.) are the saturation coefficients for the first factor. The result should be checked by finding the sum of the squares of the saturations ($\cdot8328^2 + \ldots + \cdot5392^2 = 2{\cdot}1717$). The value for the 'factor-variance' as thus determined should be identical with the value already obtained by summing the first column.

(iii) *Direct Verification (Saturation Coefficients as Weights).*—As with the perfect hierarchy (cf. Table II above) so with any table, the exact or the approximate values for the saturation coefficients may themselves be used as weights. With the appropriate divisor the method of weighted summation should then lead back to the saturations originally employed. This more direct method is especially suitable for the final verification (and minor corrections) of the figures obtained by one of the methods of successive approximation; it may also be used, without such preliminaries, when the table is nearly hierarchical or when only a rough and rapid approximation is required. In the latter case it affords a useful check on figures obtained by simple summation.

Here let us take the figures just reached in Table VI, *c*, as our weights. We proceed as for Table II above, and follow the working instructions explained on pp. 451–4. For the present set of correlations the working is shown in Table VI, *d*. The product sums, it will be remembered, have simply to be divided by the variance, which in turn is estimated by dividing the total of the product-sums (7·6741) by the total of the weights (3·536). The saturation coefficients thus obtained are, it will be seen, virtually the same as those just used for weighting.[1] The simple divisor here employed does not eliminate from the product-sums any constant factor which may have increased or decreased all the weights in equal proportion. Usually, therefore, the variance as thus estimated must be finally checked by adding the squares of the saturations. In the present example an almost complete agreement has already been secured by the final calculations in Table VI, *c*. Should the two values not agree precisely, we should have to divide by the square root of their ratio as before (cf. p. 454).

Further Factors.—To obtain saturation coefficients for the second and subsequent factors (if required) we follow the general procedure already explained. A theoretical hierarchy is constructed by multiplying the first factor-saturations by each other; the residuals are then analysed by one of the methods of weighted summation just described.

The complete set of saturation coefficients obtained by weighted summation is shown in the last three columns of Table VII. It will be seen that the sums of the squares of the saturation coefficients for each test are precisely the same as those obtained by the simple summation method, being in each case those required to complete

[1] I have carried the calculations to one more decimal place in order to facilitate the demonstration of certain equalities in Table VII. This has yielded a slight modification in the figures for the third and last tests: and incidentally the modification secures a perfect agreement in the two estimates for the variance.

the table of intercorrelations in such a way as to yield a matrix of minimal rank. Evidently, therefore, when the diagonal values have been first filled in by the full method of simple summation (after complete reiterations), there is no need for a complete reiteration of the process of weighted summation.

The totals of the bipolar factors are now no longer zero if added as they stand. But if the figures for one factor are first weighted by the corresponding saturations for any other, then the weighted totals should be zero: (two product-sums, with the figures in the first column as multipliers, are given at the foot of the Table). This means that the novel characteristic of the saturations, as corrected by weighted summation, is that *they are now entirely uncorrelated.*

Here the individual figures do not differ greatly from those obtained by simple summation. As a rule, in addition to (i) their lack of correlation, the figures obtained by weighted summation show two further related characteristics: (ii) they make the first and earlier factors responsible for rather more of the total variance (e.g. here 2·1703 instead of 2·1654) and leave less to be explained by the last factors (e.g. here 0·1458 instead of 0·1494); and (iii) within each of the earlier factors they tend to space out the differences between the several tests more plainly, usually increasing the highest saturation and diminishing the lowest. With either method, however, the correlation matrix is completely and precisely accounted for by the same minimal number of factors; here three factors only are sufficient in both cases.

The results enable us to see by concrete illustration the obvious objections to setting the total variance (or self-correlation) for each test equal to 1·00, particularly when the correlation matrix is small. (i) The proposal assumes that the total variance of a highly complex mental process (like composition) is no larger than the total variance of simpler processes (like spelling or speed of writing); whereas we should naturally expect that the variance of composition (which is not only highly complex, but actually includes the processes of spelling and quick writing) would be decidedly larger: my own method of calculation treats them as equal to ·72, ·38, and ·39 respectively. (ii) The proposal further assumes that the *specific* variances peculiar to the several tests stand in almost perfect negative correlation with their communalities or *common* variances, and therefore with their general tendency to correlate (e.g. here it implies that the specific variance of a simple test like writing would be ·61 and that of a complex test like composition only ·28): whereas we should naturally expect either no correlation at all, or, if anything, a small positive rather than a large negative correlation. (iii) It gives all these specific variances the largest possible value that is compatible with the observed intercorrelations as given in the table; i.e. it assigns a maximum amount of importance to the very factors for which there is a minimum amount of evidence. (iv) According to the principles on which the foregoing analysis

TABLE VII

SATURATION COEFFICIENTS

	A. By Simple Summation.						B. By Weighted Summation.		
	i. First Approximation.			ii. Minimum Rank.			Final Approximation.		
	1st Factor.	2nd Factor.	3rd Factor.	1st Factor.	2nd Factor.	3rd Factor.	1st Factor.	2nd Factor.	3rd Factor.
Comp.	·818	·197	·042	·822	·199	·073	·8325	·1594	·0567
Read.	·589	·490	—·163	·597	·503	—·159	·6225	·4577	—·1952
Spell.	·512	·355	·071	·506	·342	·086	·5238	·3185	·0610
Hand.	·473	—·405	—·140	·472	—·401	—·129	·4486	—·4336	—·1006
Draw.	·592	—·488	—·139	·597	—·498	—·144	·5690	—·5387	—·1126
Writ.	·549	—·139	·262	·546	—·144	·272	·5390	—·1553	·2795
	(Unweighted)						(Weighted)		
Totals	...	·010	—·067		·001	—·001	...	—·0002	·0008

is based (principles that are also those adopted by Spearman and Thurstone) the proposal would force us to assume that with each specific factor the weighting for every test except its own is exactly zero, and therefore *exactly equal*. This would be intelligible if we could assume that each of the processes tested—writing, for example—called into play a large and important specific 'ability' or 'brain-centre,' peculiar to itself and sharply localized. Actually, what is specific to each set of test-measurements, taken in isolation, with only a single application of each test, is not some unique ability or highly specialized neural process, but certain minor accidental conditions accompanying its administration or involved in its construction. Yet even these conditions can hardly be considered as remaining strictly constant for all the other tests. If they do not influence them positively, they are likely (in some small measure) to influence them negatively and in varying degrees. Thus, to be consistent, all general-factor methods of analysis (including Thurstone's centroid method) should treat the specific factor, not as a large primary ability in a 'simple structure' with a positive saturation for one test and zero saturations for the remainder, but as a small *bipolar* factor. With Hotelling's method of principal components (which also assumes the variances to be unity) there is at least one such bipolar factor which is strictly specific (in the sense of having a positive saturation for one test only); but this is mainly an artefact, since its large negative saturations are required merely to neutralize the enlarged positive saturations which appear in the preceding factors owing to the enlargement of their variances.

For special purposes yet other working methods may be tried. For

example, we might have started with the method of weighted summation, without any preliminary determination of the self-correlations by simple summation. We could then have begun by seeking the best-fitting factorial matrix of rank one; then of rank two; and so on: i.e. the first fit would contain in each diagonal cell the square of one saturation only; the second fit would contain the sum of the squares of two; and so on. This would again lead to a factorial matrix of minimum rank, with every column uncorrelated with every other. Such an approach might be of theoretical interest, but is hardly a practical procedure, since, as a rule, it leads to precisely the same figures by what in the end is not a shorter but a more cumbersome procedure.

Again, if we assume that the results of correlating persons and correlating tests should be equivalent (as they should under certain conditions, e.g. when the selection of both persons and tests is ' random '), we should expect each bipolar column of saturation coefficients to add up to zero both when unweighted and when weighted by one of the other columns. A process of successive approximation with weighted and unweighted summation used alternatively will usually yield a close approach to this result. But whatever self-correlations we assume, and whatever rank we accept, there is an obvious gain in having a factorial matrix whose columns are uncorrelated: any further linear transformations into which the correlation matrix enters—e.g. rotating axes, calculating regressions, comparing results with those of correlating persons—are greatly simplified.

If, however, the self-correlations or variances are known from the outset, then, as we have seen, successive approximation may be entirely avoided by ' table-by-table ' multiplication—a procedure that is sometimes convenient with a very small table of covariances (for illustrations, see [102], p. 185). Or again a triangular matrix of positive saturations may be obtained by the earlier procedure due to Lagrange (described, e.g., in Bôcher's *Higher Algebra*, 1907, p. 131), which will be found to fit certain correlation problems quite well.

Partial Correlations.—It should be noted that the residual correlations obtained as above by simple subtraction do not give the correlations that would be found in a sample population selected so as to be homogeneous for the factors subtracted. For example, we may desire to know what degree of correlation would exist between Reading, say, and Spelling (or Arithmetic) in a class or form where the children had been so selected as to be upon practically the same level of general ability. For this purpose the full formula for partial correlation must be employed. This is equivalent to dividing the residual correlation by the mean residual variance of the two correlated tests[1]—a process which greatly enlarges the figures. However, in

[1] [93], p. 306, and *L.C.C. Report* [35], p. 57. In my earliest article [16] this further adjustment was not applied, because the problem was to determine the *relative* importance of the first or general factor and the remainder. And in most forms of multiple-factor analysis it is, as a rule, omitted. That, however, should not obscure the fact that multiple-factor analysis (and, indeed, single-factor analysis according to Spearman's approach) is essentially a development of partial correlation: cf. *Am. J. Psychol.*, XV, 1904, p. 256 (where Yule's formula for partial correlation is cited), and Yule's determinantal solution for the partial variances ([110], pp. 267 *et seq.*).

studying these secondary factors, it is often more effective to use actual homogeneous samples, which would at once yield large correlations of this order, than to produce the effects of homogeneity by mere computation, thus continuing the analysis on hypothetical figures derived at one or two removes from a random sample.

(2) *GROUP-FACTOR METHOD*

The group-factor method is the more appropriate when the correlated variables (tests or persons) have been so selected as to form discontinuous groups. For the sake of simplicity I shall again use an artificial table to illustrate the procedure. Let us suppose that the observed correlations are as shown in the body of Table VIII. The procedure is as follows.

(i) *First Factor*

Formula : Saturation coefficient, $r_{ag} =$

$$\frac{\sum_j r_{aj}}{\sqrt{\Sigma R_{BC}}\left\{\sqrt{\frac{\Sigma R_{AB}}{\Sigma R_{AC}}}+\sqrt{\frac{\Sigma R_{AC}}{\Sigma R_{AB}}}\right\}}$$

where $\sum_j r_{aj}$ denotes the sum of the ath column in the rectangular sub-matrix $\begin{bmatrix} R_{BA} \\ R_{CA} \end{bmatrix}$ and ΣR_{BC} denotes the sum of all the correlations in the sub-matrix R_{BC}.

1. As before, it is helpful to begin by finding the total intercorrelation of each test by adding each column of observed coefficients (not shown in Table VIII).

2. With the aid of these totals, rearrange the table so as to exhibit the general trend or pattern. Start, as usual, with getting the figures as nearly as possible into a hierarchical order, by first arranging them according to the descending size of the totals. Group-factors, if present, will interrupt the steady descent of the coefficients in the rows and columns, and produce sudden rises in the correlations they affect. Bring these enlarged figures as close to the leading diagonal as possible.[1] The final result should be a series of square blocks of high correlations placed astride the leading diagonal (and perhaps overlapping) with rectangular blocks of comparatively low correlations elsewhere.

[1] If the table is too large or too irregular for the rearrangement to be made by simple inspection, we may either correlate the rows or estimate their contour by roughly graphing the figures, and then bring rows with high row-correlations, or with the same general contour, as near to one another as possible (cf. p. 300).

II. MULTIPLE-FACTOR ANALYSIS

(2) GROUP-FACTOR METHOD $\left(r_{ag}=\dfrac{\sum_j r_{aj}}{\sqrt{\Sigma R_{BC}}\left\{\sqrt{\dfrac{\Sigma R_{AB}}{\Sigma R_{AC}}}+\sqrt{\dfrac{\Sigma R_{AC}}{\Sigma R_{AB}}}\right\}}\right)$

TABLE VIII

OBSERVED INTERCORRELATIONS AND CALCULATION OF SATURATIONS FOR FIRST FACTOR

	Correlations.	Totals.	Correlations.	Totals.	Correlations.	Totals.
	.. ·75 ·65 ·75 .. ·62 ·65 ·62 ..		·54 ·45 ·36 ·48 ·40 ·32 ·42 ·35 ·28	1·35 1·20 1·05	·27 ·18 ·09 ·24 ·16 ·08 ·21 ·14 ·07	·54 ·48 ·42
Sub-totals .			1·44 1·20 ·96	3·60	·72 ·48 ·24	1·44
	·54 ·48 ·42 ·45 ·40 ·35 ·36 ·32 ·28	1·44 1·20 ·96	.. ·42 ·30 ·42 .. ·28 ·30 ·28 ..		·18 ·12 ·06 ·15 ·10 ·05 ·12 ·08 ·04	·36 ·30 ·24
Sub-totals .	1·35 1·20 1·05	3·60			·45 ·30 ·15	·90
	·27 ·24 ·21 ·18 ·16 ·14 ·09 ·08 ·07	·72 ·48 ·24	·18 ·15 ·12 ·12 ·10 ·08 ·06 ·05 ·04	·45 ·30 ·15	.. ·18 ·27 ·18 .. ·10 ·27 ·10 ..	
Sub-totals .	·54 ·48 ·42	1·44	·36 ·30 ·24	·90		
Totals .	1·89 1·68 1·47		1·80 1·50 1·20		1·17 ·78 ·39	
Divisor .	·21		·30		·39	
Sat. Coeffs.	·90 ·80 ·70		·60 ·50 ·40		·30 ·20 ·10	

3. Mark off the square blocks, as shown in the table, and prolong the lines of separation so as to partition the entire table into rectangular sub-matrices. Unless there is considerable overlapping, the lines of separation will be clearly shown by the sudden rise or fall of correlations, e.g. in rows 4 to 6 of the correlation matrix the correlations rise from the 3rd column to the 4th (instead of decreasing), and fall from the 6th to the 7th much more sharply than in rows 1 to 3. Similarly, in rows 7 to 9 they rise from the 6th column to the 7th. We may label these sub-matrices according to the usual notation for a partitioned matrix as follows :

$$R = \begin{bmatrix} R_{11} & R'_{21} & R'_{31} \\ R_{21} & R_{22} & R'_{32} \\ R_{31} & R_{32} & R_{33} \end{bmatrix}$$

The square blocks lying along the leading diagonal (R_{11}, R_{22}, R_{33}) contain correlations affected by group-factors as well as by the first factor. Hence these blocks must be omitted from all the calculations for the first factor. As shown at the head of this section, a slight but obvious complication in the primary summation formula is entailed by these omissions.

4. Add each short column in each of the oblong blocks in order to obtain its sub-total (1·35, 1·20, etc.).

5. Add the sub-totals to obtain the total for each block :

$$(\Sigma R_{21} = 3{\cdot}60,\ \Sigma R_{31} = 1{\cdot}44,\ \Sigma R_{32} = {\cdot}90).$$

6. Find the (curtailed) totals for each column (i.e. the totals obtained with the figures in the square diagonal blocks still omitted) by adding the totals of the two half-columns (1·35 + ·54 = 1·89, etc.).

7. Calculate the divisor for the curtailed totals in each double block from the square roots of the three block-totals, appropriately arranged. Thus, for the total for the first column (1·89), which is one of the three columns making up the double block $\begin{bmatrix} R_{21} \\ R_{31} \end{bmatrix}$, the divisor will be

$$\sqrt{\Sigma R_{32}}\left\{\sqrt{\frac{\Sigma R_{21}}{\Sigma R_{31}}} + \sqrt{\frac{\Sigma R_{31}}{\Sigma R_{21}}}\right\}.$$

Note that the figures inside the bracket are the totals of the two blocks to which the curtailed column belongs, and the figure outside the bracket is the total for the third remaining block. We thus obtain the following three divisors :

$$\sqrt{\cdot 90}\left\{\sqrt{\frac{3\cdot 60}{1\cdot 44}}+\sqrt{\frac{1\cdot 44}{3\cdot 60}}\right\}=\cdot 21,$$

$$\sqrt{1\cdot 44}\left\{\sqrt{\frac{3\cdot 60}{\cdot 90}}+\sqrt{\frac{\cdot 90}{3\cdot 60}}\right\}=\cdot 30,$$

$$\sqrt{3\cdot 60}\left\{\sqrt{\frac{1\cdot 44}{\cdot 90}}+\sqrt{\frac{\cdot 90}{1\cdot 44}}\right\}=\cdot 39,$$

8. Divide the curtailed totals of the columns by the appropriate divisor (1·89 ÷ ·21 = ·90, etc.).

The result will be the saturation coefficients for the first factor.

(ii) *Second Factor*

Formulæ: Residual correlation, $r'_{ab} = r_{ab} - r_{ag}\, r_{bg}$.

$$\text{Saturation coefficient, } r'_{as} = \frac{\sum_i r'_{ai}}{\sqrt{\sum_j \sum_i r'_{ji}}}.$$

1. Multiply each of the saturation coefficients by each of the others, thus obtaining a perfect hierarchy of theoretical correlations attributable to the first or general factor (·90 × ·80 = ·72, ·90 × ·70 = ·63, etc.). The multiplication proceeds as before, and its results are given in the lower of the paired rows in Table IX.

2. Enter the observed figures above them, and subtract the theoretical figures from the observed, as shown in Table IX (·75 − ·72 = ·03, ·65 − ·63 = ·02, etc.). When (as here) group-factors are discontinuous and do not overlap, the residuals thus obtained will be approximately zero in each cell of the rectangular blocks governed by the general factor alone; but in the square blocks along the leading diagonal, where the effects of a group-factor are superimposed on those of the general factor, there will be positive residuals due to the operation of the former.

3. Take each square block of residuals in turn, treat it as a separate hierarchy, insert estimated values for the diagonal coefficients, and analyse the whole by the method described in Section IA. In Table X the central block of residuals from Table IX is factorized as an example.

The four sets of saturation coefficients obtained by this twofold procedure are set out in Table XI. It will be seen that they account perfectly for the 'observed' correlations. To render the illustration more definite these correlations were artificially constructed at the outset from these very figures. And it will be noted that the group-

TABLE IX

Hierarchy for First Factor and Calculation of Residuals

Sat. Coeff.	·90	·80	·70	·60	·50	·40	·30	·20	·10
Obs. Corr.	..	·75	·65	·54	·45	·36	·27	·18	·09
·90	..	·72	·63	·54	·45	·36	·27	·18	·09
Residuals	..	·03	·02	·00	·00	·00	·00	·00	·00
Obs. Corr.		..	·62	·48	·40	·32	·24	·16	·08
·80		..	·56	·48	·40	·32	·24	·16	·08
Residuals	·03	..	·06	·00	·00	·00	·00	·00	·00
Obs. Corr.				·42	·35	·28	·21	·14	·07
·70				·42	·35	·28	·21	·14	·07
Residuals	·02	·02	..	·00	·00	·00	·00	·00	·00
Obs. Corr.				..	·42	·30	·18	·12	·06
·60				..	·30	·24	·18	·12	·06
Residuals				..	·12	·06	·00	·00	·00
Obs. Corr.				..	..	·28	·15	·10	·05
·50					..	·20	·15	·10	·05
Residuals				·12	..	·08	·00	·00	·00
Obs. Corr.				..	..	..	·12	·08	·04
·40				..	..	..	·12	·08	·04
Residuals				·06	·08	..	·00	·00	·00
Obs. Corr.							..	·18	·27
·30							..	·06	·03
Residuals							..	·12	·24
Obs. Corr.							..	..	·10
·20							..	..	·02
Residuals							·12	..	·08
Obs. Corr.									..
·10									..
Residuals							·24	·08	..

TABLE X

ONE SUB-MATRIX OF RESIDUALS AND CALCULATION OF SATURATION COEFFICIENTS FOR 2ND GROUP-FACTOR

	Residuals.			Totals.
	[·09]	·12	·06	·27
	·12	[·16]	·08	·36
	·06	·08	[·04]	·18
Total	·27	·36	·18	$·81 = ·90^2$
Sat. Coeff. . . .	·30	·40	·20	·90
Squares	·09	·16	·04	—

factor method thus leads back to the original saturations from which the table was constructed.

Unless we are to regard Thurstone's requirements as excluding *a priori* anything like a general factor, we may say that the factor-pattern obtained fits his description of a 'simple structure': all the coefficients are positive, and each test and each factor (except the general) has a maximum number of zero saturations.

Note.—The above formula is not valid when the correlation matrix is partitioned into *more* than 9 sub-matrices, i.e. when there are more than three different rectangular blocks influenced by the first or general factor only. In that case, instead of omitting the square diagonal blocks, R_{11}, R_{22}, etc., we have first to estimate their expected totals. This is done by treating the remaining totals ΣR_{21}, ΣR_{31}, . . . ΣR_{23}, etc., as forming a hierarchy with the entries in the leading diagonal unknown. Applying the simple-summation formula in its original form, we can deduce appropriate values for ΣR_{11}, ΣR_{22}, . . . etc., and thence the grand total for the entire hierarchy, $\Sigma\Sigma_{pq} R_{pq}$ say. The formula for the saturation coefficients is then as follows:

$$r_{ag} = \frac{\sum_j r_{aj} \sqrt{\sum_p \sum_q R_{pq}}}{\sum_p \sum_q R_{pq} - \Sigma(R_{1a} + R_{2a} + \ldots + R_{fa})}$$

where j as before extends over all the tests correlated except those affected by the same group-factor as a, R_{1a}, . . . R_{fa} denote the f sub-matrices in which the ath test falls (including the square diagonal blocks), and p and q extend over all the f sub-matrices without exception.

TABLE XI

SATURATION COEFFICIENTS OBTAINED BY GROUP-FACTOR AND GENERAL-FACTOR METHODS

(C) Group-factor Method.				General-factor Method.					
				(A) By Simple Summation.			(B) By Weighted Summation.		
General Factor	Group-Factors.			General Factor.	Bipolar Factors.		General Factor.	Bipolar Factors.	
(i)	(ii)	(iii)	(iv)	(i)	(ii)	(iii)	(i)	(ii)	(iii)
·90	·00	·00	·10	·860	—·199	—·173	·886	—·071	—·124
·80	·00	·00	·30	·791	—·200	—·238	·822	—·080	—·197
·70	·00	·00	·20	·684	—·168	—·181	·709	—·064	—·147
·60	·00	·30	·00	·621	—·174	·181	·630	—·080	·213
·50	·00	·40	·00	·546	—·168	·290	·549	—·086	·312
·40	·00	·20	·00	·414	—·116	·121	·420	—·054	·143
·30	·60	·00	·00	·433	·512	·000	·352	·570	·018
·20	·20	·00	·00	·238	·150	—·015	·214	·184	·005
·10	·40	·00	·00	·195	·363	·015	·140	·388	·022
Totals: Positive				Unweighted:	+1·025	+·607	Weighted:	+2·943	+3·760
Totals: Negative					—1·025	—·607		—2·943	—3·760
Totals: Algebraic					0·000	0·000		0·000	0·000

Thus, with the above table, we first estimate and check the values for ΣR_{11}, ΣR_{22}, . . ., $\Sigma\Sigma R_{pq}$, as follows :

TABLE XII

	[5·76] 3·60 1·44	3·60 [2·25] ·90	1·44 ·90 [·36]	10·80 6·75 2·70
Total . .	10·80	6·75	2·70	$20·25 = 4·5^2$
Quotient .	2·40	1·50	·60	
Square . .	5·76	2·25	·36	

The saturation coefficient for (say) the first test is then :

$$r_{1g} = \frac{1·89\sqrt{20·25}}{20·25 - (5·76 + 3·60 + 1·44)}$$
$$= ·90 \text{ as before.}$$

(3) *GENERAL-FACTOR METHOD*

Suppose now that the discontinuities in the tables of observed correlations had not seemed quite so obvious, and that we had decided to apply the more usual ' general-factor method ' instead : what results should we have obtained ?

I shall leave the reader to carry out, as an exercise for himself, the detailed calculations according to the instructions given above.[1] He will find satisfactory results are reached if he fills the diagonal with the values for the 'communalities' supplied by the group-factor analysis. Table XI, A and B, give the final figures, viz. the saturations obtained by simple and by weighted summation respectively.

As before, the peculiarity of the simple-summation method is that the totals of the bipolar saturations add up to exactly zero, and of the weighted summation method that the bipolar columns are uncorrelated both with each other and with the column for the

[1] The only novel feature arises in the treatment of the residuals obtained after the first or general factor has been eliminated. In this case it is impossible, by means of negative weighting, to convert *all* the negative residuals into positive. As before, the choice of tests to be negatively weighted is determined by the general pattern, and checked by the signs of the resulting saturation coefficients. Here the computer who followed the usual rules alone (reversing tests with most negatives or largest absolute totals) might be tempted to give the 5th test an opposite sign to the others in the first group of 6, and perhaps the 8th an opposite sign to the others in the last group of 3. But this would evidently spoil the pattern. It may be that such results as this partly explain why those who follow Thurstone's modification of the summation method have supposed that its results must be meaningless until rotated.

general factor, so that the *weighted* totals add up to zero.[1] With the former method the figures fit the observed two-place correlations with discrepancies of less than ·01 ; with the latter they yield a fit that is even closer still. Thus the factor pattern furnished by general-factor methods tends to be more economical or parsimonious than the ' simple structure ' furnished by the group-factor method or an equivalent rotation. The reason is clear. Where the group-factor method yields a factor with 6 zeros, the general-factor method will substitute a factor with 6 negative figures (as well as the 3 positive figures appearing in both), so that this latter is helping to do the work of a second group-factor; and the negative figures of the two bipolar factors will between them dispense with the need for a third group-factor.

It is evident that the two main forms of analysis lead to virtually the same conclusions. The results may be summed up as follows: (1) All the nine tests are influenced by a general factor whose importance decreases from one test to another in a definite order, the order being the same with all three methods. (2) The nine tests can be divided into three groups of three, each group being influenced positively by a different factor from the rest, the division being the same with all three methods: thus, with each method we see (i) that the *last* three tests are influenced positively by a large special factor which does not enter positively into the first six ; and (ii) the *middle* three tests are influenced by a smaller special factor which does not enter positively into the first three and does not enter at all into the last three (or hardly at all) ; moreover, since we may reverse the signs for the last bipolar factor, we may add (iii) the first *three* tests are in effect influenced positively by a still smaller special factor, which does not influence the middle three in the same way and does not enter at all into the last three (or hardly at all). Finally, (3) within each group of three the influence of its special factor diminishes from one test to another in an order which is the same with all three methods. The chief difference between the results of the group-factor method and those of the two general-factor methods is that the former assumes that a non-positive influence is identical with a complete absence of influence, the latter assumes that it may be a negative influence and may vary in amount.

Now, if we prefer the former method of expression, but start with a summation method of analysis, we can obviously translate our bipolar general factors into group-factors by ' rotation of axes.' Thus, if we started (as Thurstone would do) by the simple summation method, we could ' rotate ' the factors shown in Table XIA,

[1] With ' weighting ' defined as on p. 456, this holds of both methods.

and from them we could obtain the figures shown in Table XIc, approximately if we followed his 'graphic' procedure,[1] exactly if we used a more adequate method and retained enough factors and decimals. If we started with weighted summation, the rotation could still be carried out, though Thurstone's criticism of such methods seems to imply that it could not: indeed, the necessary calculations would be much easier, owing to the zero correlations between the rows.[2] But in either case it is evident that we save a vast amount of labour, and probably reach more precise figures, if, when we want a 'simple structure,' we seek it by the direct 'group-factor method' according to the procedure described above.

Of the three main methods described in this Appendix I recommend *simple summation for general purposes, weighted summation for exact research, and the group-factor method for special cases of discontinuous distribution.* The student, however, who is new to factorial work should practise all three. Let him begin with a small fictitious table of unit figures only (cf. next Appendix); and then try his hand with physical measurements before proceeding to mental. Let him take, say, half a dozen body measurements for half a dozen persons (e.g. sitting height, length of leg, and of arm, girth of chest, waist, and hips for a homogeneous and for a dichotomous group, i.e. mixed males and females, or pyknics and leptosomics), and factorize covariances and correlations for both traits and persons by every procedure available, including 'simple averaging' (p. 396). Physical measurements present much the same problems as mental; and the interpretations are easier to visualize.[3] The theoretical advantages of weighted summation will quickly become obvious. Like Thurstone's method, it yields a factor-pattern of minimal rank; like Hotelling's, it yields the best possible fit. It would thus seem to combine the merits of both.

[1] Thurstone in his rotations would presumably begin by seeking the factors containing zero saturations (which with him are usually *overlapping* group-factors) and leave the general factor (i.e. the factor with no negative or zero saturations) to emerge, if at all, only when the group-factors do not account for the correlation.

[2] The method has already been described above (p. 304); and therefore need not be repeated here.

[3] The idea that there is a 'contrast' between results from physical and mental measurements (the latter alone yielding hierarchies) and that neither 'bodily dimensions' nor 'assessments of physical maturity' show signs of general factors (Spearman [56], p. 142, Thomson [132], p. 279) is not borne out by my own data. Both with adults and with children the factor-variances follow the same rough progression as is obtained with psychological tests (Table VII, p. 358). For Spearman's own chief example ([56], p. 141) v_i/n = ·49, ·12, ·09: (the two group-factors—covering the 3 head measurements and the 2 skull-breadth measurements—were only to be expected). Of course, if we mix linear with non-linear measurements (e.g. weight) and anatomical with physiological ([56], p. 144), the hierarchical tendency is bound to be obscured.

APPENDIX II

ANALYSIS OF A MATRIX INTO ITS LATENT ROOTS AND LATENT VECTORS

THE ultimate result of analysing a correlation table by the method of 'least squares' or 'weighted summation' is to reduce the initial matrix of measurements to terms of the latent roots and latent vectors of the correlation matrix. It can thus be expressed as the product of three component matrices—a diagonal matrix, pre-multiplied and post-multiplied by two semi-orthogonal matrices, or, in matrix notation, as $M = LV^{\frac{1}{2}}P$, where L is the 'modal matrix' of 'latent vectors,' V the diagonal matrix of 'latent roots' or factor variances, and P the semi-orthogonal 'population' matrix of factor-measurements. Where it is necessary to distinguish matrices obtained by correlating tests and persons respectively, a subscript t or p will be affixed.

As I have stated in the text, this mode of analysis seems to me the most logical and the most useful; and, since the nomenclature of matrix algebra may be unfamiliar, the following tables are appended to illustrate the simplicity of the procedure. Incidentally they will serve to demonstrate that, provided the initial matrix is suitably standardized, the resulting 'factors' are the same, whether we begin by covariating persons or tests.

In order that the reader can follow the working mentally, I have taken a small fictitious table of integers, and have calculated covariances, instead of reducing the figures to standard measure and so calculating correlations. If desired, the correlations and the factor-saturations[1] could at once be obtained by dividing the covariances and the factor-loadings by the square roots of the variances of the initial measurements: e.g. in Table II, by $\sqrt{56}$, $\sqrt{20}$, and $\sqrt{20}$.

The initial matrix, M, is set out at the head of Table I, and is supposed to give the marks of 4 persons (p_1, p_2, p_3, p_4) in three

[1] These will not represent the same set of factors as would have been obtained by analysing the correlations directly. If we required factor-saturations directly deduced from the correlations rather than covariances, we should have to start by normalizing the initial measurements at the outset. That would involve working with decimal fractions running into several figures, and would be too complicated to follow mentally.

tests (t_1, t_2, t_3). It is standardized so that the averages for all persons and all tests are zero.

The preliminary analysis proceeds according to the method described in the foregoing Appendix (pp. 459 f.). Thus, in 'correlating tests' we first calculate the covariance matrix, $MM' = R_t$, shown at the top of Table II(*a*), and then obtain the factor-loadings for the first and second factors, $F_tF'_t = R_t$. The sum of the

TABLE I

To obtain the factor-measurements (P_t) by means of the variances (V) and factor-loadings (F_t): $P_t = WM = V^{-1}F'_tM$.

		M				
		p_1	p_2	p_3	p_4	
		−6	2	0	4	t_1
		3	1	−1	−3	t_2
		3	−3	1	−1	t_3
	F'_t t_1 t_2 t_3					
f_1	$2\sqrt{14}$ $-\sqrt{14}$ $-\sqrt{14}$	$-18\sqrt{14}$	$6\sqrt{14}$	0	$12\sqrt{14}$	F'_tM
f_2	0 $\sqrt{6}$ $-\sqrt{6}$	0	$4\sqrt{6}$	$-2\sqrt{6}$	$2\sqrt{6}$	
	V^{-1}					
f_1	$\frac{1}{84}$ 0	$-\frac{3}{\sqrt{14}}$	$\frac{1}{\sqrt{14}}$	0	$\frac{2}{\sqrt{14}}$	P_t
f_2	0 $\frac{1}{12}$	0	$\frac{2}{\sqrt{6}}$	$-\frac{1}{\sqrt{6}}$	$\frac{1}{\sqrt{6}}$	
	f_1 f_2	p_1	p_2	p_3	p_4	

Note.—To find the product of two matrices, proceed as in multiplying two determinants, i.e. multiply the elements of the *rows* of the pre-factor by the corresponding elements of the *columns* of the post-factor, and add. Thus, taking the first row of F' and the first column of M,

$$(2\sqrt{14}) \times (-6) + (-\sqrt{14}) \times (3) + (-\sqrt{14}) \times (3) = -18\sqrt{14}$$

The dash affixed to a symbol, e.g. F', denotes that the original matrix, F, has been 'transposed,' i.e. rewritten with the original rows as columns.

squares of the factor-loadings then gives the factor-variance for that factor. (See Table II(*a*), where the factor-loadings and variances for the first factor are calculated explicitly.) Here, as the reader can easily test for himself, simple summation happens to yield the same result as weighted summation.

The factor-measurements for each person, P_t, can be computed by simply adding his test-measurements, after first weighting them by the factor-loading for the test concerned (F'_tM): the totals are then usually normalized, i.e. reduced to unitary standard measure, by dividing by the respective factor-variances (V) (the detailed working is shown in Table I). It will be observed that the calculation is very much simpler than the laborious procedure required in calculating regression coefficients to obtain estimated factor-measurements in the ordinary way (described by Thurstone [84], pp. 226 f., and Thomson [132], pp. 93 f.).

The latent vectors, L_t, are obtained by simply normalizing the columns of the factor-loadings, F_t: i.e. dividing each column by the root of its squares.

When 'correlating tests,' therefore, we first obtain the covariance matrix R_t; on factorizing this by the 'least-squares' method we obtain the factorial matrix, F_t; and from F_t we obtain (i) the normalized factor-loadings for tests, L_t, (ii) the factor-variances, V_t, and (iii) the factor-measurements for persons, P_t.

Table II(*b*) shows how the initial set of measurements can be reconstructed by multiplying these three component matrices.

Similarly, when 'correlating persons' we first obtain the covariance matrix, R_p; on factorizing this we obtain the factorial matrix, F_p; and from F_p we obtain (i) the normalized factor-loadings for persons, L_p, (ii) the factor-variances, V_p, and (iii) the factor-measurements for tests, P_p.

Table III(*b*) shows how the initial set of measurements can be reconstructed by multiplying these three component matrices.

It will be observed (i) that the factor-variances are the same in either case ($V_p = V_t$), and (ii) that the factor-measurements for persons obtained by 'covariating tests' are identical with the normalized factor-loadings for persons obtained by 'covariating persons,' and (iii) that the factor-measurements for tests obtained by 'covariating persons' are identical with the normalized factor-loadings for tests obtained by 'covariating tests': in short, with a measurement matrix thus standardized and thus factorized, the results are the same whether we correlate (or rather covariate) persons or tests.

TABLE II

Results of Factorizing Covariances between Tests: $M = L_t V^{\frac{1}{2}} P_t$.

(*a*) Covariance Matrix $MM' = R_t = F_t F'_t$.

	R_t			
	t_1	t_2	t_3	Total*
t_1	56	−28	−28	112
t_2	−28	20	8	56
t_3	−28	8	20	56
Total*	112	56	56	$224 = (4\sqrt{14})^2$
Total ÷ $\sqrt{224}$	$2\sqrt{14}$	$(-)\sqrt{14}$	$(-)\sqrt{14}$	0
Square	56	14	14	84 = 1st factor-variance

* As usual, before calculating the total, the tests that are to receive a negative factor-loading (here tests 2 and 3) have their signs reversed.

(*b*) Reconstruction.

$$L_t \times V^{\frac{1}{2}}{}_t \times P_t$$

$$\begin{array}{c} \\ t_1 \\ t_2 \\ t_3 \end{array}\!\!\begin{array}{c} \overbrace{}^{f_1}\ \overbrace{}^{f_2} \\ \begin{bmatrix} \frac{2}{\sqrt{6}} & 0 \\ -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \end{array} \times \begin{array}{c} \overbrace{}^{f_1}\ \overbrace{}^{f_2} \\ \begin{bmatrix} \sqrt{84} & 0 \\ 0 & \sqrt{12} \end{bmatrix} \end{array} \times \begin{array}{c} \overbrace{}^{p_1}\ \overbrace{}^{p_2}\ \overbrace{}^{p_3}\ \overbrace{}^{p_4} \\ \begin{bmatrix} -\frac{3}{\sqrt{14}} & \frac{1}{\sqrt{14}} & 0 & \frac{2}{\sqrt{14}} \\ 0 & \frac{2}{\sqrt{6}} & -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}} \end{bmatrix} \end{array} \begin{array}{l} \\ f_1 \\ f_2 \end{array}$$

$$= \begin{array}{c} \\ t_1 \\ t_2 \\ t_3 \end{array}\!\!\begin{array}{c} \overbrace{}^{f_1}\ \overbrace{}^{f_2} \\ \begin{bmatrix} 2\sqrt{14} & 0 \\ -\sqrt{14} & \sqrt{6} \\ -\sqrt{14} & -\sqrt{6} \end{bmatrix} \end{array} \times \begin{array}{c} \overbrace{}^{p_1}\ \overbrace{}^{p_2}\ \overbrace{}^{p_3}\ \overbrace{}^{p_4} \\ \begin{bmatrix} -\frac{3}{\sqrt{14}} & \frac{1}{\sqrt{14}} & 0 & \frac{2}{\sqrt{14}} \\ 0 & \frac{2}{\sqrt{6}} & -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}} \end{bmatrix} \end{array} \begin{array}{l} \\ f_1 \\ f_2 \end{array}$$

$$F^t \times P$$

M_t

	p_1	p_2	p_3	p_4
$= t_1$	−6	2	0	4
t_2	3	1	−1	−3
t_3	3	−3	1	−1

Analysis by Selective Operators: p. 494. (*a*) *Canonical Expansion of Covariance Matrix* (see p. 164).—The method of multiple-factor analysis described in the text is based on the assumption that any symmetric covariance (or correlation) matrix can be expressed as the sum of a number of perfect hierarchies, each of which consists of a ' unit hierarchy' (E_j) multiplied by the corresponding factor-variance (v_j). Such a series I call the ' canonical expansion ' of the covariance (or correlation) matrix. A unit hierarchy is formed by post-multiplying a latent vector (one column of L_t) by its transpose (the same figures written as a row), according to the usual method of constructing a theoretical hierarchy. Thus, in Table IV*a*, E_1 is formed from the first column of L_t (shown in Table II*b*), E_2 from the second column, and E_3 from the otiose column required to turn the semi-orthogonal L_t matrix into a completely orthogonal square matrix. From the weighted sum of these unit hierarchies we can reconstruct the original covariance matrix, R_t, as shown in the table. Each of the unit hierarchies represents the effect of a single pure and independent factor; and the corresponding weighted hierarchy represents the amount of correlation or covariance attributable to that factor.

(*b*) *Factorial Expansion of Measurement Matrix* (see pp. 261–5).—We can determine the contribution of such a factor to the empirical matrix of test-measurements by simply pre-multiplying that matrix by the appropriate unit hierarchy (see Table IV*b*). With the multipliers calculated in this way, the result is to yield, with E_1, the best approximation to M that is obtainable with a single factor only, namely, $M_1 = E_1 M$; with E_2, the best approximation to the residuals, $(M - M_1)$, that is obtainable with a single factor only; and so on (the closeness of the approximation being judged by the principle of least squares). Thus, if we assume that there can be only one true factor determining the initial test-measurements, and that the remainder consists of errors of measurement, then M_1 gives the best estimate of the hypothetical values. And the same result will be reached whether we start by correlating rows or

columns, persons or tests: I leave the reader to calculate unit hierarchies from Table II*b*, and so verify this statement.

The sum of all such contributions should obviously yield the original measurement matrix (as shown at the foot of Table IV*b*). If P_t were already known, we could obtain the same detailed and total result by first multiplying the rows of factor-measurements by the appropriate factor-loadings, and then adding the products, according to the familiar equation $M = F_t P_t$ (cf. Table II*b* above). But since P_t is not given and M is, it seems more logical to exhibit the direct analysis of M than its resynthesis from P_t.

Such a series of unit hierarchies is analogous to what is termed in quantum theory a 'spectral set of selective operators': each 'selects' the contribution of its factor, and so 'analyses a mixed aggregate into its pure constituents.' Thus, the mental performance measured by a given test is in effect conceived as the sum of a number of contributory reactions, mixed and heterogeneous: the selective operators sort these reactions into a few mutually exclusive classes, such as would popularly be attributed to distinct elementary

TABLE III

Results of Factorizing Covariances between PERSONS: $M = L_p V^{\frac{1}{2}} P_p$

(*a*) Covariance matrix $M' M = R_p = F_p F'_p$.

	R_p				Total*
	p_1	p_2	p_3	p_4	
p_1	54	−18	0	−36	108
p_2	−18	14	−4	8	36
p_3	0	−4	2	2	0
p_4	−36	8	2	26	72
Total*	108	36	0	72	$216 = (6\sqrt{6})^2$
Total ÷ $\sqrt{216}$	$(-)3\sqrt{6}$	$\sqrt{6}$	0	$2\sqrt{6}$	0
Squares	54	6	0	24	84 = 1st factor-variance

* As usual, before calculating the total, the persons that are to receive a negative factor-loading (here person 1) have their signs reversed.

(*b*) Reconstruction.

$$
\begin{array}{c}
(L_p = P'_i) \\
\begin{array}{c} \\ p_1 \\ p_2 \\ p_3 \\ p_4 \end{array}
\begin{array}{c} f_1 \quad f_2 \\ \begin{bmatrix} -\frac{3}{\sqrt{14}} & 0 \\ \frac{1}{\sqrt{14}} & \frac{2}{\sqrt{6}} \\ 0 & -\frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{14}} & \frac{1}{\sqrt{6}} \end{bmatrix} \end{array}
\end{array}
\times
\begin{array}{c}
(V\mathfrak{i}_p = V\mathfrak{i}_i) \\
\begin{array}{c} f_1 \quad f_2 \\ \begin{bmatrix} \sqrt{84} & 0 \\ 0 & \sqrt{12} \end{bmatrix} \end{array}
\end{array}
\times
\begin{array}{c}
(P_p = L'_i) \\
\begin{array}{c} t_1 \quad t_2 \quad t_3 \\ \begin{bmatrix} \frac{2}{\sqrt{6}} & -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}} \\ 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \end{array} \begin{array}{c} f_1 \\ f_2 \end{array}
\end{array}
$$

$$
= \overbrace{\begin{array}{c} \\ p_1 \\ p_2 \\ p_3 \\ p_4 \end{array}
\begin{array}{c} f_1 \quad f_2 \\ \begin{bmatrix} -3\sqrt{6} & 0 \\ \sqrt{6} & 2\sqrt{2} \\ 0 & -\sqrt{2} \\ 2\sqrt{6} & -\sqrt{2} \end{bmatrix} \end{array}}^{F_p}
\times
\overbrace{\begin{array}{c} t_1 \quad t_2 \quad t_3 \\ \begin{bmatrix} \frac{2}{\sqrt{6}} & -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}} \\ 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \end{array} \begin{array}{c} f_1 \\ f_2 \end{array}}^{P_p}
$$

$$
= \overbrace{\begin{array}{c} \\ p_1 \\ p_2 \\ p_3 \\ p_4 \end{array}
\begin{array}{c} t_1 \quad t_2 \quad t_3 \\ \begin{bmatrix} -6 & 3 & 3 \\ 2 & 1 & -3 \\ 0 & -1 & 1 \\ 4 & -3 & -1 \end{bmatrix} \end{array}}^{M'}
$$

'abilities.' (Note the somewhat questionable assumption, clearly brought out by the table, that for the same test each testee employs precisely the same abilities in precisely the same proportions.)

The unit hierarchies have a number of peculiar properties, which render this mode of analysis particularly convenient for mathematical manipulation and which the reader can easily verify in the present instance: they are 'idempotent,' i.e. $E_j^m = E_j$; they are mutually orthogonal, i.e. $E_i E_j = 0$; their sum is equal to the unit matrix, i.e. $E_1 + \ldots + E_n = 1$; and the 'trace' of each (sum of diagonal elements) $= 1$. (Cf. [115], 'The Unit Hierarchy and Its Properties.')

TABLE IV

Factorial Expansion of Covariance and Measurement Matrices.

(*a*) Analysis of Covariance Matrix into Weighted Sum of Unit Hierarchies: $R_t = H_1 + H_2 + \ldots + H_n = v_1E_1 + v_2E_2 + \ldots + v_nE_n$.

$$R_t = v_1 \times E_1 + v_2 \times E_2 + v_3 \times E_3$$

$$\begin{bmatrix} 56 & -28 & -28 \\ -28 & 20 & 8 \\ -28 & 8 & 20 \end{bmatrix} = 84 \begin{bmatrix} \frac{4}{6} & -\frac{2}{6} & -\frac{2}{6} \\ -\frac{2}{6} & \frac{1}{6} & \frac{1}{6} \\ -\frac{2}{6} & \frac{1}{6} & \frac{1}{6} \end{bmatrix} + 12 \begin{bmatrix} 0 & 0 & 0 \\ 0 & \frac{1}{2} & -\frac{1}{2} \\ 0 & -\frac{1}{2} & \frac{1}{2} \end{bmatrix} + 0 \begin{bmatrix} \frac{2}{6} & \frac{2}{6} & \frac{2}{6} \\ \frac{2}{6} & \frac{2}{6} & \frac{2}{6} \\ \frac{2}{6} & \frac{2}{6} & \frac{2}{6} \end{bmatrix}$$

(*b*) Analysis of Measurement Matrix into the Sum of the Contributions of its Factors: $M = E_1M + E_2M + \ldots + E_nM$.

M

	p_1	p_2	p_3	p_4	
	−6	2	0	4	t_1
	3	1	−1	−3	t_2
	3	−3	1	−1	t_3

Initial matrix.

E_1

$\frac{4}{6}$	$-\frac{2}{6}$	$-\frac{2}{6}$	−6	2	0	4	t_1
$-\frac{2}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	3	−1	0	−2	t_2
$-\frac{2}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	3	−1	0	−2	t_3

Contribution of 1st factor.

E_2

0	0	0	0	0	0	0	t_1
0	$\frac{1}{2}$	$-\frac{1}{2}$	0	2	−1	−1	t_2
0	$-\frac{1}{2}$	$\frac{1}{2}$	0	−2	1	1	t_3

Contribution of 2nd factor.

E_3

$\frac{2}{6}$	$\frac{2}{6}$	$\frac{2}{6}$	0	0	0	0	t_1
$\frac{2}{6}$	$\frac{2}{6}$	$\frac{2}{6}$	0	0	0	0	t_2
$\frac{2}{6}$	$\frac{2}{6}$	$\frac{2}{6}$	0	0	0	0	t_3

Contribution of 3rd factor

Total	−6	2	0	4	t_1
	3	1	−1	−3	t_2
	3	−3	1	−1	t_3

Reconstruction of initial matrix.

REFERENCES

[1] Lagrange, J. L. (1759), *Œuvres*, I., pp. 1–20.

[2] Gauss, C. F. (1823), *Werke*, IV, pp. 26–55.

[3] Sylvester, J. J. (1851), *Collected Works*, I, pp. 221 *et seq.*; IV, pp. 110 *et seq.*

[4] Bravais, A. (1846), 'Analyse mathématique sur les probabilités des erreurs de situation d'un point,' *Acad. des Sciences: Mémoires présentés par divers savants*, II^e Série, IX, pp. 255–73.

[5] Galton, F. (1888), 'Correlations and their Measurement,' *Proc. Roy. Soc.*, XLV, pp. 136–45.

[6] Cattell, J. McK. (1890), 'Mental Tests and Measurements,' *Mind*, XV, pp. 373–80.

[7] Pearson, K. (1896), 'Regression, Heredity, and Panmixia,' *Phil. Trans. Roy. Soc.*, Ser. A, CLXXXVII, pp. 253–318.

[8] Yule, G. U. (1897), 'On the Significance of Bravais' Formulæ for Regression, etc., in Skew Correlation,' *Proc. Roy. Soc.*, LX, pp. 477–89.

[9] Wissler, C. (with J. M. Cattell and R. C. Farrand) (1901), 'The Correlation of Mental and Physical Tests,' *Psych. Rev. Mon. Supp.*, III.

[10] Pearson, K. (1902), 'On the Influence of Natural Selection on the Variability and Correlation of Organs,' *Phil. Trans. Roy. Soc.*, CC, A, pp. 1–66.

[11] — (1904), 'Inheritance of Mental and Moral Characters in Man,' *Biometrika*, III, pp. 131–90.

[12] Spearman, C. (1904), 'General Intelligence Objectively Determined and Measured,' *Am. J. Psych.*, XV, pp. 201–93.

[13] Krueger, F., and Spearman, C. (1906), 'Die Korrelation zwischen verschiedenen geistigen Leistungsfähigkeiten,' *Z. f. Psych.*, XLIV, pp. 50–114.

[14] Weldon, W. F. R. (1906), 'Inheritance in Animals and Plants,' *ap.* Strong, T. B. (ed.) *Lectures on the Method of Science.*

[15] Bôcher, M. (1907), *Introduction to Higher Algebra* (New York: Macmillan).

[16] Burt, C. (1909), 'Experimental Tests of General Intelligence,' *Brit. J. Psych.*, III, pp. 94–177.

[17] Burt, C. (1909), 'The Experimental Study of General Intelligence,' *Child Study*, IV, pp. 78–100.

[18] Thorndike, E. L. *et al.* (1909), 'The Relation of Accuracy in Sensory Discrimination to General Intelligence,' *Am. J. Psych.*, XX, pp. 364–9.

[19] Brown, W. (1910), 'Some Experimental Results in the Correlation of Mental Abilities,' *Brit. J. Psych.*, III, pp. 296–322.

[20] Burt, C. (1911), 'Experimental Tests of Higher Mental Processes and their Relation to General Intelligence,' *J. Exp. Ped.*, I, pp. 93–112.

[21] Stern, W. (1911), *Die Differentielle Psychologie* (Leipzig: Barth).

[22] Burt, C. (1912), 'The Inheritance of Mental Characters,' *Eugenics Review*, IV, pp. 1–33.

[23] Burt, C., and Moore, R. C. (1912), 'The Mental Differences between the Sexes,' *J. Exp. Ped.*, I, pp. 273–84, 233–88.

[24] Spearman, C., and Hart, B. (1912), 'General Ability; its Existence and Nature,' *Brit. J. Psych.*, V, pp. 51–84.

[25] Yule, G. U. (1912), *Introduction to the Theory of Statistics* (London: Griffin).

[26] Cullis, C. E. (1913, 1918, 1925), *Matrices and Determinoids*, Vols. I, II, III (Cambridge: University Press).

[27] Burt, C. (1914), 'The Measurement of Intelligence by the Binet Tests,' *Eugenics Review*, VI, pp. 36, 140 *et seq.*

[28] Spearman, C. (1914), 'The Theory of Two Factors,' *Psych. Rev.*, XXI, pp. 101–15.

[29] Burt, C. (1914–29), *Reports of the Psychologist to the London County Council.*

[30] —— (1915), 'The General and Specific Factors underlying the Primary Emotions,' *Brit. Ass. Ann. Rep.*, pp. 694–6.

[31] Lankes, W. (1915), 'Perseveration,' *Brit. J. Psych.*, VII, pp. 387–398.

[32] Webb, E. (1915), 'Character and Intelligence,' *Brit. J. Psych., Mon. Supp.* No. 3.

[33] Bickersteth, M. E., and Burt, C. (1916), Report of Results of Mental and Scholastic Tests. *Rep. IVth Ann. Conf. Educ. Ass.*, pp. 30–37.

[34] Thomson, G. H. (1916). 'A Hierarchy without a General Factor,' *Brit. J. Psych.*, VIII, pp. 271–81.

[35] Burt, C. (1917), *Three Reports on the Distribution and Relations of Educational Abilities* (London: P. S. King).

[36] Garnett, J. C. M. (1919), ' On Certain Independent Factors in Mental Measurement,' *Proc. Roy. Soc.*, A. XCVI, pp. 102–5.

[37] —— (1919), ' General Ability, Cleverness and Purpose,' *Brit. J. Psych.*, IX, 345–66.

[38] Thomson, G. H. (1919), ' On the Cause of Hierarchical Order among Correlation Coefficients,' *Proc. Roy. Soc.*, A, XCV, pp. 400–408.

[39] Brown, W., and Thomson, G. (1920), *Essentials of Mental Measurement* (Cambridge : University Press).

[40] Thomson, G. (1920), ' General Versus Group Factors in Mental Activities,' *Psych. Rev.*, XXVII, pp. 173–190.

[41] Burt, C. (1921), *Three Reports on Mental and Scholastic Tests* (London : P. S. King).

[42] Aveling, F., and Hargreaves, H. L. (1921), ' Suggestibility in Children,' *Brit. J. Psych.*, VIII, pp. 1–15.

[43] Keynes, J. M. (1921), *A Treatise on Probability* (London : Macmillan).

[44] Sheppard, W. F. (1923), *From Determinant to Tensor* (Oxford : Clarendon Press).

[45] Spearman, C. (1923), *The Nature of Intelligence and the Principles of Cognition* (London : Macmillan).

[46] Burt, C. (1923), ' The Mental Differences between Individuals,' Pres. Address, Psychology Section, *Brit. Ass. Ann. Rep.*, pp. 215–39.

[47] Kelley, T. L. (1923), *Statistical Method.* (New York : Macmillan).

[48] Burt, C. (1924), *ap.* Board of Education : *Report of Consultative Committee on Psychological Tests of Educable Capacity* (London : H.M. Stationery Office).

[49] Thorndike, E. L., and Others (1925), *The Measurement of Intelligence* (New York : Columbia University Press).

[50] Fisher, R. A. (1925–34), *Statistical Methods for Research Workers* (Edinburgh : Oliver & Boyd).

[51] Burt, C. (1925), ' A Statistical Study of Junior County Scholarship Examinations,' *L.C.C. Reports.*

[52] Spearman, C., and Holzinger, K. J. (1925), ' Note on the Sampling Error of Tetrad Differences,' *Brit. J. Psych.*, pp. 86–8.

[53] Burt, C., Gaw, F., Spielman, W., Smith, May, *et al.* (1926), *A Study in Vocational Guidance* (London : H.M. Stationery Office).

[54] Davey, C. M. (1926), 'Group, Verbal and Pictorial Tests of Intelligence,' *Brit. J. Psych.*, XXVI, pp. 1–20.

[55] Thomson, G. H., and Bailes, S. (1926), 'The Reliability of Essay Marks,' *Forum of Education*, IV, pp. 85–96.

[56] Spearman, C. (1927), *The Abilities of Man: Their Nature and Measurement* (London: Macmillan).

[57] Thomson, G. H. (1927), 'The Tetrad Difference Criterion,' *Brit. J. Psych.*, XVII, pp. 235–55.

[58] Hargreaves, H. L. (1927), 'The Faculty of Imagination,' *Brit. J. Psych. Mon. Supp.*, No. 10.

[59] Pearson, K., and Moul, M. (1927), 'The Mathematics of Intelligence. I. Sampling Errors in the Theory of a Generalized Factor,' *Biometrika*, XIX, pp. 246–91.

[60] Kelley, T. L. (1928), *Cross Roads in the Mind of Man* (Stanford University Press).

[61] Dodd, S. C. (1928), 'The Theory of Factors,' *Psych. Rev.*, XXXV, pp. 211–34, 261–79.

[62] Howland, R. (1928), *Phil. Mag.* (*Ser.* vii), VI, pp. 839–42.

[63] Spearman, C. (1928), 'Pearson's Contribution to the Theory of Two Factors,' *Brit. J. Psych.*, pp. 95–101.

[64] Wilson, E. B. (1928), 'On Hierarchical Correlation Systems,' *Proc. Nat. Acad. Sc.*, XIV, pp. 283–91.

[65] Turnbull, H. W. (1928), *The Theory of Determinants, Matrices and Invariants* (London: Blackie).

[66] Mackie, J. (1928), 'The Sampling Theory as a Variant of the Two-Factor Theory,' *J. Educ. Psych.*, XIX, pp. 614–21.

[67] Spearman, C. (1930), 'G and After: a School to end Schools,' *ap. Psychologies of* 1930, pp. 339–66.

[68] Holzinger, K. J. (1930), *Statistical Résumé of the Spearman Two-Factor Theory* (Chicago: Univ. Chicago Press).

[69] Burt, C. (1931), *ap.* Report of Consultative Committee of the Board of Education on *The Primary School*, Appendix III (London: H.M. Stationery Office).

[70] Stephenson, W. (1931), 'Tetrad Difference for Verbal Sub-Tests relative to Non-Verbal Sub-Tests,' *J. Educ. Psych.*, XXII, pp. 334–50.

[71] Weyl, H. (1931), *The Theory of Groups and Quantum Mechanics* (London: Methuen).

[72] Tryon, R. C. (1932), 'So-called Group Factors as Determiners of Ability,' *Psych. Rev.*, XXXIX, pp. 403–39.

[73] Turnbull, H. W., and Aitken, A. C. (1932), *Introduction to the Theory of Canonical Matrices* (London: Blackie).

[74] MacDuffee, C. C. (1933), 'The Theory of Matrices,' *Ergebn. d. Math.*, V, No. 5.

[75] Burt, C. (1933), 'The Psychology of Art,' *ap.* Burt, Jones, *et al.*, *How the Mind Works*, pp. 294 *et seq.*

[76] Brown, W. (1933), 'The Mathematical and Experimental Evidence for the Existence of a Central Intellective Factor (g),' *Brit. J. Psych.*, XXIII, pp. 171–9.

[77] Brown, W., and Stephenson, W. (1933), 'A Test of the Theory of Two Factors,' *Brit. J. Psych.*, XXIII, pp. 352–71.

[78] Irwin, J. O. (1933), 'A Critical Discussion of the Single Factor Theory,' *Brit. J. Psych.*, XXII, pp. 371–81.

[79] Hotelling, H. (1933), 'Analysis of a Complex of Statistical Variables into Principal Components,' *J. Educ. Psych.*, XXIX, pp. 417–41, 498–520.

[80] Beebe-Center, J. (1933), *Pleasantness and Unpleasantness* (New York: Century Press).

[81] Wedderburn, J. H. M. (1934), *Lectures on Matrices* (with bibliography of 549 references) (New York: Amer. Mathematical Society).

[82] Alexander, W. P. (1935), 'Intelligence, Concrete and Abstract,' *Brit. J. Psych. Mon. Supp.*, No. 19.

[83] Holzinger, K. J. (1935), *Preliminary Reports on Spearman-Holzinger Unitary Trait Study*, Nos. 5 and 6.

[84] Thurstone, L. L. (1935), *The Vectors of the Mind* (Chicago: University of Chicago Press).

[85] Kelley, T. L. (1935), *Essential Traits of Mental Life* (Harvard: University Press).

[86] Tryon, R. C. (1935), 'A Theory of Psychological Components —an Alternative to Mathematical Factors,' *Psych. Rev.*, XLII, pp. 425–54.

[87] Thomson, G. H. (1935), 'On Complete Families of Correlation Coefficients and their Tendency to Zero Tetrad Differences: including a Statement of the Sampling Theory of Abilities,' *Brit. J. Psych.*, XXVI, pp. 63–92.

[88] —— (1935), 'The Definition and Measurement of G,' *J. Educ. Psych.*, XXVI, pp. 241–62.

[89] —— (1935), 'The Factorial Analysis of Human Abilities,' *The Human Factor*, IX, pp. 180–5.

[90] Simmins, C. A. (1935), 'Studies in Experimental Psychiatry: Deterioration of G in Psychotic Patients,' *J. Mental Science*, XX, pp. 1–29.

[91] Stephenson, W. *et al.* (1935), 'The Spearman Factors in Psychiatry,' *Brit. J. Med. Psych.*, XIV, pp. 100 *et seq.*

[92] Stephenson, W. (1935), 'Correlating Persons instead of Tests,' *Character and Personality*, IV, pp. 17–24.
[93] Burt, C. (1936), 'The Analysis of Examination Marks: a Review of Methods of Factor Analysis in Psychology,' *ap.* Hartog, P., and Rhodes, E. C., *Marks of Examiners* and separate reprint (London: Macmillan & Co.).
[94] Emmett, W. G. (1936), 'Sampling Error and the Two Factor Theory,' *Brit. J. Psych.*, XXVI, pp. 362–87.
[95] Philpott, S. J. F. (1936), 'The Meaning of G and other Factors,' *Brit. J. Psych.*, XXVII, pp. 196–224.
[96] Stephenson, W. (1936), 'The Foundations of Psychometry,' *Psychometrika*, I, pp. 195–210.
[97] —— (1936), 'The Inverted Factor Technique,' *Brit. J. Psych.*, XXVI, pp. 344–61.
[98] ——(1936), 'Introduction to Inverted Factor Analysis,' *J. Educ. Psych.*, XXVII, pp. 353–67.
[99] Thomson, G. H. (1936), 'Some Points of Mathematical Technique in the Factorial Analysis of Ability,' *J. Educ. Psych.*, XXVII, pp. 37–54.
[100] —— (1936), 'Boundary Conditions in the Common Factor Space,' *Psychometrika*, I, pp. 155–64.
[101] Burt, C. (1937), 'Correlations between Persons,' *Brit. J. Psych.*, XXVIII, pp. 59–95.
[102] —— (1937), 'Methods of Factor-Analysis with and without Successive Approximation,' *Brit. J. Educ. Psych.*, VII, pp. 172–95.
[103] Ledermann, W. (1937), 'On the Rank of the Reduced Correlational Matrix in Multiple-Factor Analysis,' *Psychometrika*, II, pp. 85–93.
[104] Cattell, R. B. (1937), 'Measurement versus Intuition in Applied Psychology,' *Character and Personality*, pp. 114–131.
[105] Vernon, P. E. (1937), 'The Stanford-Binet Test as a Psychometric Method,' *ibid.*, pp. 99–113.
[106] Holzinger, K. J. (1937), *Student Manual of Factor Analysis* (Chicago: Department of Education).
[107] —— and Swineford, F. (1937), 'The Bi-factor Method,' *Psychometrika*, II, pp. 41–54.
[108] Thurstone, L. L. (1937), *Psychometrika*, II, pp. 73–6.
[109] Fisher, R. A. (1937), *The Design of Experiments* (Edinburgh: Oliver & Boyd).
[110] Yule, G. U., and Kendall, M. G. (1937), *Introduction to the Theory of Statistics* (11th ed.) (London: Griffin).

[111] Aitken, A. C. (1937), 'The Evaluation of the Latent Roots and Vectors of a Matrix,' *Proc. Roy. Soc. Edin.*, LVII, pp. 269–304.

[112] Spearman, C. (1938), 'The Proposed Explanation of Individual Differences of Ability by Sampling,' *Brit. J. Psych.*, XXIX, pp. 182–91.

[113] —— (1938), *Psychology Down the Ages* (London: Macmillan).

[114] Burt, C. (1938), 'The Analysis of Temperament,' *Brit. J. Med. Psych.*, XVII, pp. 158–88.

[115] —— (1938), 'The Unit Hierarchy and its Properties,' *Psychometrika*, III, pp. 151–68.

[116] —— (1938), 'Factor-Analysis by Submatrices,' *J. of Psych.*, VI, pp. 339–75.

[117] Myers, C. S., Spearman, C., Thomson, G. H., Burt, C., *et al.* (1938), 'Discussion on Quantitative Methods in Psychology' (abstracts only), *Proc. Roy. Soc.*, Ser. B, CXXV, pp. 415–34.

[118] Dewar, H. (1938), 'A Comparison of Tests of Artistic Appreciation,' *Brit. J. Psych.*, VIII, pp. 29–49.

[119] Williams, E. D., Winter, L., and Woods, J. M. (1938), 'Tests of Literary Appreciation,' *Brit. J. Educ. Psych.*, VIII, pp. 265–84.

[120] Vernon, P. E. (1938), 'The Assessment of Psychological Qualities by Verbal Methods' (London: H.M. Stationery Office).

[121] Burt, C. (1938), 'Recent Developments of Statistical Method in Psychology,' *Occupational Psychology*, XII, pp. 169–77.

[122] Thurstone, L. L. (1938), *Primary Mental Abilities* (Chicago: University of Chicago Press).

[123] Buros, O. K. (1938), *The* 1938 *Mental Measurements Yearbook.* (Rutgers Univ. Pub.).

[124] Thomson, G. H. (1938), 'The Influence of Univariate Selection on the Factorial Analysis of Ability,' *Brit. J. Psych.*, XXVIII, pp. 451–9.

[125] Thomson, G. H., and Lederman, W. (1938), 'The Influence of Multivariate Selection on the Factorial Analysis of Ability,' *Brit. J. Psych.*, XXIX, pp. 66–73.

[126] Mather, K. (1938), *The Measurement of Linkage in Heredity* (London: Methuen).

[127] Frazer, R. A., Duncan, W. J., and Collar, A. R. (1938), *Elementary Matrices and some Applications to Dynamics* (Cambridge: University Press).

[128] Burt, C. (1939), 'The Relations of Educational Abilities,' *Brit. J. Educ. Psych.*, IX, pp. 45–71.

[129] Burt, C. (1939), 'The Factorial Analysis of Emotional Traits,' *Character and Personality*, VII, pp. 238–54, 285–99.
[130] Davies, M. (1939), 'The General Factor in Correlations between Persons,' *Brit. J. Psych.*, XXIX, pp. 404–21.
[131] Tryon, R. C. (1939), *Cluster Analysis* (Chicago: University Press).
[132] Thomson, G. H. (1939). *The Factorial Analysis of Human Ability* (London: University of London Press).
[133] Eysenck, H. (1939), 'Critical Notice of "Primary Mental Abilities,"' *Brit. J. Educ. Psych.*, IX, pp. 270–5.
[134] Cast, B. M. D. (1939), 'Efficiency of Different Methods of Marking,' *ibid.*, IX, pp. 257–69, X, pp. 49–60.
[135] Vernon, P. E. (1939), 'Educational Abilities of Training College Students,' *ibid.*, pp. 233–50.
[136] Stephenson, W. (1939), 'Two Contributions to the Theory of Mental Testing,' *Brit. J. Psych.*, XXX, pp. 19–35, 230–47.
[137] Thomson, G., Spearman, C., Stephenson, W., and Burt, C. (1939), Symposium on 'The Factorial Analysis of Ability,' *ibid.*, pp. 71–108.
[138] Burt, C., and Stephenson, W. (1939), 'Alternative Views on Correlations between Persons,' *Psychometrika*, IV, pp. 269–82.
[139] Vernon, P. E. (1939), *The Measurement of Abilities* (London: University of London Press).

Note.—The foregoing references mainly cover work that is easily accessible, and easily intelligible, to the English student. The more recent issues of the American journal *Psychometrika* contain articles of special interest to the advanced mathematical student; but unfortunately—particularly at the present moment—they are not readily obtainable in this country. Technical articles on factor-analysis have also appeared in recent numbers of the *Proceedings of the Royal Society of Edinburgh*. The relevant German work on matrix-algebra should be familiar to senior research workers who propose to study the mathematical aspects: the more important references are summarized in [74] and [81]. For the beginner the best mathematical introduction is [84] chap. I. He should then proceed to [132]. A revision of his school knowledge of determinants and an occasional reference to the earlier chapters of [15], [26], [65], or [73] will be helpful in elucidating the commoner terms and theorems.

INDEX OF AUTHORS

Aitken, A. C., 255, 299, 305, 469, 498, 501
Alexander, W. P., 7, 11, 212, 236, 267, 302, 499
Aristotle, 34, 105, 111, 140, 156, 232

Bacon, F., 39, 97
Bartlett, M. S., 159, 225
Bartlett, R. J., 117
Binet, A., 55–6, 58, 172, 175, 217, 267, 410
Bocher, M., 307, 336, 476, 495
Bowley, A. L., 45, 52, 66
Bravais, A., 52, 66
Broad, C. D., 35, 47, 65, 222, 234
Brown, W., 5, 353, 496, 497

Campbell, N. F., 118, 121
Cast, B. M. D., 271, 279, 339, 502
Cattell, J. McK., 4, 133, 271, 495
Cattell, R. B., 56, 67, 285, 413, 500
Center, J. B., 191, 203, 393, 401
Collar, A. R., 82, 255, 469, 501
Cullis, C. E., 149, 299, 336, 496

Darbishire, A. D., 52
Davey, C. M., 142, 339, 498
Davies, M., 157, 187, 191, 352, 502
Dewar, H., 130, 178, 184, 501
Duncan, W. J., 82, 255, 469, 501

Eddington, A., x, 213, 220, 226, 236, 256

Fisher, R. A., x, 40, 68, 77, 180, 182, 272–8, 289, 436, 497, 500
Frazer, R. A., 82, 255, 469, 501

Galton, F., 4, 6, 133, 495
Garnett, J. C. M., 88, 151, 216, 257, 412, 497
Guilford, J. P., 14, 128, 145, 158, 189, 212, 340, 411, 422

Hargreaves, H. L., 412, 498
Hill, J. C., 54, 230
Holzinger, K. J., 212, 255, 268, 298, 499, 500
Hotelling, H., 73, 255, 286, 320, 469, 499

Irwin, J. O., 145, 499

Jeffreys, H., 23, 47, 240
Johnson, W. E., x, 28, 35, 45, 115, 233, 240, 266
Joseph, H. W. B., 39, 105, 112
Jung, C. G., 410, 418, 440

Kelley, T. L., 11, 68, 73, 105, 158, 212, 255, 320–7, 498, 499
Keynes, J. M., x, 35, 38, 66, 68, 222, 497
Kretschmer, E., 113, 244, 410, 416, 440
Krüger, F., 495

Leibniz, G. W., 125
Lexis, W., 40

McDougall, W., x, 176, 217, 388, 406, 410
Mill, J. S., 3, 31, 33, 39, 45, 69, 221, 224
Moore, R. C., 284, 421, 496

Pearson, K., 4, 52, 81, 128, 219, 340, 437, 495
Philpott, S. J. F., 500
Poincaré, H., 25, 117
Porphyry, 111, 114

Quetelet, A., 113

Reyburn, H. A., 60, 107
Rhodes, E. C., 172, 178
Russell, B., 45, 93, 118, 125, 180, 220, 238

Seymour, W. D., 278, 339
Simmins, C. A., 59, 192, 499
Snedecor, G. W., 40, 42, 274
Spearman, C., v, x, 3, 40, 42, 142, 155, 211, 257, 287–8, 297, 332, 343, 359, 412, 419–20, 495–502
Spielman, W., 177, 497
Stebbing, L. S., 115
Stephenson, W., xi, 74, 116, 177–201, 268, 280, 296, 378–84, 411–19, 500, 502
Studman, L., 192, 413, 417, 499

Temple, G., 100
Thomson, G. H., x, 5, 10, 52, 65, 72, 159, 200, 211–17, 269, 291–4, 373, 448, 496–502
Thorndike, E. L., 4, 225, 335, 430
Thouless, R. H., 240
Thurstone, L. L., xi, 7, 11, 22, 72, 106, 157, 161, 172, 210, 262, 297, 311–18, 349, 358–64, 455
Tippett, L. H. C., 68
Titchener, E. B., 83, 223, 240
Tryon, R. C., 268–9, 499, 502

Vernon, P. E., 223, 285, 501

Webb, E., 7, 88, 192, 363
Weldon, W. F. R., 52, 495
Williams, E. D., 196, 501
Wilson, J. Cook, 39
Woods, J. M., 130, 196, 501
Woodworth, R. S., 83
Wundt, W., 83, 217

Young, G. W., 85, 115, 180
Yule, G. U., 4, 44, 50, 53, 148, 152, 180, 275, 289, 353, 496, 500

INDEX OF SUBJECTS

Abilities, 3, 6, 51, 67, 105, 203, 236
  æsthetic, 76, 178
  arithmetical, 5, 59, 76, 203, 311
  educational, 5, 29, 55, 126, 230, 311–17, 338, 460–76
  mechanical, 76, 142, 315
  motor, 60, 203, 315
  primary, 6, 84, 211, 237, 251, 311
  verbal, 6, 18, 55, 74–6, 142, 203, 311, 407
Addition, postulates of, 130–2
Æsthetic appreciation, 76, 121, 178, 192–6
Agreement, method of, 31, 33
Analogy, 28–39
Analysis of variance, 19, 40, 66, 180, 235, 248, 271–80, 365, 433
Analytic method, 61, 121
Asthenic type, *see* Types
Atomic uniformity, principle of, 65, 222
Attributes, 46, 65, 99, 110, 115, 250

Backwardness, 15, 19, 54, 58, 96, 243
Bi-factor method, 268, 298
Binet tests, 55–9, 172, 175, 217
Bipolar factor, *see* Factors

Canonical expansion, 165, 265, 323, 326, 491
Causation, 4, 38, 65–71, 86, 218, 228, 231
  axiom of, 31, 69, 220, 228
Centroid method, 72, 225, 326, 385, 395; *see also* Summation method, simple
Chance, *see* Factors, chance
Classification, 19, 77, 202, 250, 309
Cluster analysis, 268–9
Colour equations, 20, 84, 120
Common elements, 52, 66
Communalities, 109, 157, 160, 328, 332, 341, 467
Components, 4, 64, 84, 211, 257
  principal, 26, 255, 327, 467
  resolution of, 78, 84
Composition, English, 76, 204, 271, 461
Concomitant variation, 3, 108
Connotation, 110, 160, 200
Correlation, 4, 6, 52, 68, 258, 289, 294
  intra-class, 275–7, 346
  of persons, x, 7, 109, 124, 169–209, 289–94, 390–3, 420
  ratio, 274–8, 346
Cosine law, 60, 88, 91, 164
Covariance, 6, 149, 234, 258, 280–8
Cyclic overlap, 143, 297, 461

Degrees of freedom, 274
Denotation, 110, 160, 200
Diagonal entries, 152, 328, 331, 341, 392, 448, 450, 460, 466–7, 482
Direction cosines, 109, 153, 321
Distensive magnitude, *see* Measurement

Economics, factor-analysis in, ix, 133, 248
Economy, postulate of, *see* Parsimony, law of
  ratio, 347
Educational guidance, v, 15, 54–9, 227–30, 243
Eduction, 75–9, 65
Emotions, 7, 75, 83, 176, 192, 388
Empirical method, 61

Energy, mental, 67, 82, 86–92, 217, 236, 251
Error, *see* Factors, chance ; Sampling, random
Extension, *see* Denotation
Extensive magnitude, *see* Measurement
Extraversion, 76, 193, 407, 410–17, 428, 434, 443

Factors :
as abilities, 199, 211, 236
as averages, 74, 198, 249, 332, 395
as axes of coordinates, 79–84
as causes, 65–71, 211, 218–21, 229
as patterns, 64, 75–8, 237, 418
as principles of classification, 94–100, 202–4, 212, 249
as specifying relations, 204, 227, 241
as statistical abstractions, 213, 254, 408
definition of, 4, 210–13, 256–9
kinds of :
bipolar, 24, 76, 102, 113, 134, 175, 309, 390, 454–9
central, *see* general
chance, 39, 102, 114, 139, 144
common, 56, 66, 102
dominant, 102, *see also* general
error, *see* chance
fractional, 187, 267, 318, 413
general, 59, 102, 113, 139, 158, 161, 168, 194, 225, 289, 297 ; *see also* Intelligence, General emotionality, General factor for persons
generating, 65
genetic, 8 ; *see also* Genotypes ; Inheritance, Mendelian
group, 16, 59, 72, 103, 139, 187, 225, 392
highest common, 107–8
individual, *see* specific
non-fractional, 187, 267, 413
oblique, 40, 264–8
orthogonal, 79, 83, 88, 260
overlapping, 45, 57, 143, 297
Factors (*continued*)
kinds of (*continued*)
particular, *see* group
positive, *see* dominant
random, *see* chance
singular, *see* specific
specific, 55, 102, 111, 139, 141, 154–8, 329
type, *see* Types
unique, *see* specific
universal, *see* general
verbal, *see* Abilities, verbal
logical nature of, 95–138
metaphysical nature of, 12, 210–48
physical nature of, 12, 160, 217
reification of, 72, 213–18
Factor-equation, 72, 252
Factor-loadings, 6, 73, 149, 290
Factor-measurements, 72–4, 115, 136, 235, 290, 307–10, 329–32, 401
Factor-saturations, 6, 73, 115, 235, 299, 322, 393–9
negative, 236, 296, 309, 318, 455 ; *see also* Factors, bipolar
Factor-theories :
dual-factor theory, 140–1, 156, 162
four-factor theory, 101–4, 109, 139, 249, 298
inverted-factor theory, 169–209
multiple-factor theory, 161–7, 184
sampling theory, 52, 65, 159–61, 209
single-factor theory, 143–6
three-factor theory, 26, 139, 296
two-factor theory, 5, 141–3, 160, 184, 296, 335, 447
Faculties, 55, 213, 218, 232, 335
Fluency, 192, 412
Forces, mental, 52, 218
resolution of, 52, 84–92, 234
Four-factor theorem, 103, 166
Four-factor theory, 101–4, 109, 139, 249, 298
Fourier's theorem, 165, 211

General emotionality, 7, 96, 406, 413, 419

General intelligence, *see* Intelligence
General factor for persons, 171, 194, 340, 420
General-factor methods, 295–319, 365, 459–77, 484–6
General factors, definition of, 102
Genetic factors, *see* Factors
Genotypes, 9, 114, 377, 385
Gestalt, 67, 77, 180, 224, 232, 238, 251
Group-factor methods, 156, 295–319, 365, 477–83
Group-factors, definition of, 102
Groups, theory of, 93, 242

Halo effect, 196, 421
Hereditary differences, *see* Inheritance
Hierarchical criterion, 54, 159, 337–64
  theorem of added, 164, 265, 323, 326, 491
Hierarchy, 24, 27, 97, 150–4, 337–64
  definition of, 149; *see also* Matrix, rank of
  of specific intelligences, 4, 144
  unit, *see* Unit hierarchy
Highest common factor, 107–8
Homogeneity, criterion of, 50, 147

Independence, 53, 83, 147, 257, 261; *see also* Factors, orthogonal
  criterion of, 50, 147
Individual factors, *see* Specific factors
Induction, 21–44, 46, 221–4
Inheritance, 8, 47, 57, 141, 230, 408, 441
  Mendelian, 8, 246, 408, 440
Instincts, 96, 134, 243, 388
Intelligence, 10, 21–3, 57, 74, 96, 195, 213, 217, 314
Intensive magnitude, *see* Measurement
Interaction, 278–9
Intercolumnar criterion, 149, 300, 335, 342, 346
Introversion, 76, 407, 410–17, 428, 434, 443
Inverted-factor theory, *see* Q-technique
Item-analysis, 58

Key qualities, 63, 75–6, 96

Latent roots, 92, 153, 166, 263, 320, 487
Latent vectors, 92, 153, 166, 263, 320, 487
Least-noticeable differences, 130, 134, 282
Least squares, method of, 26, 148, 164, 263, 327, 385; *see also* Summation method, weighted
Linear relations, 119, 241, 258

Matrix algebra, 79, 82, 143, 238–9, 258–65, 487–94
  compartite, 301
  doubly centred, 292, 382, 457
  factorial, 73, 261, 330, 486
  modal, 486; *see also* Latent vectors
  multiplication of, 260, 488
  population, 73, 262, 330, 486
  rank of, 107, 146, 263, 454, 459
  rank one, 149; *see also* Hierarchy
  root of, *see* Latent roots
Measurement, 115–138
  distensive, 125, 133
  extensive, 118, 124
  intensive, 118
  postulates of, 115–19, 130–1
  units of, 6, 129, 134, 173, 281
Memory, 6, 16, 55, 59
Mental deficiency, 55, 95, 217
Mental energy, *see* Energy
Multiple-factor theorem, 164
Multiple-factor theory, 161–6, 184, 459–486
Multiplicative theorem, 47

Natural kinds, 35, 60, 69, 105, 224
Neurodynamics, 91, 217

Operations, 127, 256

Operators, 100, 227; *see also* Selective operators
Overlapping factors, *see* Factors

Parsimony, law of, 14, 27, 106, 108, 329
Percentiles, 123
Perseveration, 6, 192, 412
Personalistics, 182
Phenotypes, 9, 114, 377, 385
Physical measurements, 112–14, 174
  types, *see* Types
Point, definition of, 85
Population, *see* Sampling; Universe
Postulates:
  a priori, 25, 39, 42–4
  monotonic, 131
  of addition, 130–2
  of atomic uniformity, 65, 223
  of causation, 4, 69
  of economy, 14, 27, 106, 108, 329
  of independence, 52, 83
  of limited independent variety, 224
  of measurement, 119, 130–1
  of parsimony, *see* of economy
  of simplicity, 23–5, 240
Predicables, 110–14, 250
Prediction, 20–64, 228–31
Primary abilities, *see* Abilities
Primary colours, 84, 122
Probability, 34, 47, 220, 229, 240
Product theorem, 47, 49–50, 150, 447
Progressive delimitation, 64, 106
Property, *see* Predicables
Proportionality criterion, 54, 148
Proprium, *see* Predicables
Psychoanalysis, 229, 244
Psychodynamics, 91
Psychogram, 63, 78, 91, 371, 426, 443
Psychophysical methods, 135
P-technique, *see* Correlations between persons
Pyknic type, *see* Types.

Q-technique, 169–206, 386, 394, 418
Quantum theory, 100, 150, 162, 225, 467

Random errors, *see* Factors, chance
Random sampling, *see* Sampling, random
Randomization, 40, 278–9
Rank, definition of, 146
  of correlation matrix, 107, 263, 454, 459
  of measurement matrix, 263
Reciprocity principle, 175, 282, 291–4, 406, 442, 487–93
Reflection of signs, 76, 458, 465
Regression, 48, 68, 78, 262, 329
Relations, 33, 89, 118, 227, 234–9
Representative sampling, *see* Sampling
Reversibility, principle of, 205
Rotation of axes, 81, 236, 259, 303–17, 334, 361, 485
R-technique, 177, 185–8

Sampling, 6, 145
  random, 6, 36, 44, 101
  representative, 36, 145, 247
  simple, *see* random
  theory, 52, 65, 159–61, 209
Saturation coefficients, *see* Factor-saturations
Scales of measurement, 126–7
  standard, 128, 183
Selection, 66
Selective operators, 100, 264–5, 318, 323, 491–4
Sex differences, 8, 142
  linkage, 8, 266
Significance, 6, 187, 338–40
Simple structure, 23, 27, 240, 297, 302, 335
  summation, *see* Summation
Single-factor theorem, 146
Single-factor theory, 54, 143–5, 449
Singular factors, *see* Factors
Specific abilities, *see* Abilities
  factors, *see* Factors
    definition of, 102
Spectral analysis, 100, 165, 267
Spherical coordinates, 81
  trigonometry, 81, 88

Summation method, simple, 27, 107, 152, 320–32, 365, 449–51, 461–6
weighted, 27, 91, 152, 320–32, 365, 451–4, 467–76
Symmetry criterion, 41, 156, 281

Temperament, 83, 192, 371, 387–443
Tetrad difference criterion, 54, 149, 335–7, 353, 357
Theorem of added hierarchies, 164, 265, 323, 326, 491
four-factor, 103, 166
multiple-factor, 163, 184
single-factor, 146
three-factor, 138
two-factor, 141–3, 180, 184, 383
Theory of groups, *see* Groups, theory of
Three-factor theory, 26, 139, 296
Trace, 341
Trace-ratios, 341, 350–5, 368
Transitivity, postulate of, 119
Trigonometrical method, 255, 320
Trigonometry, spherical, 81, 88
Two-factor theory, 4, 24, 101, 184
Type factor, *see* Factors
Types, 36, 63, 138, 207, 224, 400
distribution of, 429–36, 443
imagery, 170, 421
measurement of, 424–8
physical, 112–13, 174, 245
racial, 113, 246
social, 247
temperamental, 83, 189, 371, 393–402, 424–43
Typology, 188, 418

Unit hierarchy, 64, 153, 164, 264, 323
Units of measurement, *see* Measurement
Universe, 179, 289

Variable, 46, 92, 99, 115, 179–82, 241
Variance, 19, 93, 153, 235, 263, 283–6, 392; *see also* Analysis of variance
Vectors, 78, 84, 147; *see also* Latent vectors
Vocational guidance, 55, 63, 229
psychology, 60, 171
selection, 60–5

Weber-Fechner law, 134
Weighted summation, *see* Summation