

represented at the Seventh International Congress of Genetics be requested to discuss the teaching of statistics in relation to genetics and to transmit their final resolutions to the next congress.

In connexion with the Congress, a meeting of mouse geneticists was held on August 24 to consider a set of nomenclature rules drawn up by a recently formed Committee on Mouse Genetics Nomenclature and to discuss details of a Mouse Genetics News Service. The director and staff of the Roscoe B. Jackson Memorial Laboratory in Bar Harbor, Maine, U.S.A., have offered facilities for the publication in mimeographed form of the *Mouse Genetics News*. It was suggested that a register of stocks and the various pure lines should be drawn up; this should end the confusion in the naming of pure lines used in various laboratories which has arisen during the last few years. Stock lists of all the laboratories concerned should be

published from time to time. It was further suggested that notice should be given by a laboratory before any stocks are discontinued; it has happened several times in the past that valuable material has been irretrievably lost, because every laboratory has relied on other places for its maintenance. The News Service should also arrange for exchange of stocks, and it is hoped that its activities may be extended to rabbits and other rodents. The meeting also discussed the establishment of centres, preferably in the United States, for the maintenance and safe keeping of stocks, particularly of genes (pathological and otherwise) which are not purposely kept by the fancy. It was urged that this matter should receive the immediate consideration of the News Service, and that an appeal should be made to the Carnegie and Rockefeller Foundations for financial assistance. F. A. E. CREW.

MEASUREMENT IN PSYCHOLOGY*

BY R. J. BARTLETT

IN the opening paragraph of "Psychology Down the Ages", Prof. C. Spearman concludes a list of difficulties facing any who seek to define psychology, with the questions: "Do the data at its disposal include what can properly be called 'measurements'? Is it, or can it ever hope to be, or should it as much as try to be, a systematic science at all?" Dr. William Brown, in the realm of medical psychology, answers the first question in the negative. Forced into metaphysics, he writes: "Determinism, although a postulate for psychology, cannot be accepted as anything proved. In physical science there is empirical proof of it to a certain extent through measurement. By measurement we can prove to a certain extent, within certain limits of error, the conservation of mass, the conservation of energy. . . . But there is no measurement of that sort possible in psychology. . . . The observations of psychology are primarily qualitative, not quantitative." But the writer of "Essentials of Mental Measurement" could not leave it there. He added a footnote: "But mental measurement, in a derived form, is possible in the domain of mental tests and of the psycho-physical methods."

The conflict between philosophy and experimental psychology, thus exemplified, is of long standing. Malebranche argued that mental states could be distinguished qualitatively, but not

quantitatively. Leibniz declared that mathematical treatment was impossible in psychology as it lacked continuous magnitudes. Kant said that psychology possessed no mathematical magnitudes for the reason that, "whereas the phenomena of matter possess two variables, space and time, those of mind have only the single one, time". For these and many later philosophers psychology possesses neither measurable phenomenon nor the means to measure. Still worse! Some physicists seem to have reached the same conclusion. For them measurement is impossible in psychology.

Notwithstanding criticism from philosophy and physics, the word measurement is deeply imbedded in the literature of psychology, and of late years its use has been greatly on the increase. Experimental psychology is little more than fifty years old. Galton's "Inquiries into Human Faculty" appeared in 1883, Wundt went to Leipzig in 1875, Weber died in 1878 and Fechner in 1887. Since then an ever-growing body of workers has been applying experimental method to psychological problems and developing its own methods of measurement and mathematical treatment of data obtained. Contemplating this work against the declaration that measurement is impossible in psychology, Prof. Spearman comments: "But the path of science is paved with achievements of the allegedly unachievable. And in point of fact, mathematical treatment is perhaps just the region where psychology has made its steadiest

* From the presidential address to Section J (Psychology) of the British Association, delivered in Dundee on September 1.

and most surprising advances." My task is to attempt a review of experimental psychology that will reveal the senses in which the word measurement is being used by psychologists.

PSYCHO-PHYSICAL MEASUREMENTS

When we turn to actual measurements being made in psychological laboratories, we find in use metre scales and foot rules, balances and weights, chronoscopes and stop watches, resistance boxes, galvanometers, photometers and any other measuring instrument that ancient wisdom or modern physics has made available and happens to be of service in the task in hand.

One general group of tasks consists in measuring the results of bodily activity carried out under controlled conditions. Take as an example the product of a person's behaviour when asked to draw lines on blank foolscap paper of length equal to half the width of the paper. In one experiment 2,048 such lines were drawn and measured. It is probable that no two lines were exactly the same length, and that measurement to 0.01 mm. would have proved this, while obscuring a law binding them together.

Choosing a coarser unit, we measure to the nearest millimetre and find 92 lines of 103 mm. length at the modal value of results scattered from 69 mm. to 135 mm. Relative to the material our unit is still fine, so by addition we find the numbers of lines that would have fallen into each group had we measured with units 2, 3, 4, 5 and 6 mm. respectively. With increase of the size of unit the numbers approximate more and more closely to a smooth curve. Further, as we take out the figures for measurement with a unit of 6 mm., we recognize in them an approximation to the terms of the binomial expansion $(1+1)^n$. Then, using the slightly smaller unit of $5\frac{2}{3}$ mm., we obtain a very close approximation to the binomial point curve. For a particular case we have established the fact that by suitable choice of unit our experimental data fall into groups closely approximating to the binomial distribution of their number, 2048. This is but a particular case of a general law.

The Gaussian continuous function, commonly called the *normal curve* or *curve of error*, in the form

$$y = \frac{N}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2}$$

gives close approximations to the values of binomial distributions, and is usually taken to represent the law of distribution of the results of the bodily expression of mental intention. The frequency with which results of this nature are obtained in

psychological experiments leads to the conclusion that an appreciable scatter is an essential feature of our data, and that, while the old-time physicist or chemist can rest satisfied with the mean of two determinations, we must make at least fifty before we can be reasonably sure of the values of our constants. Experimental psychology could have made little progress without accepting what Dr. Darwin, last year, in his presidential address to Section A, called "the fuzziness inherent in absolutely all facts of the world", when he was arguing "that the subject of probability ought to play an enormously greater part in our mathematical-physical education".

Kant reached the conclusion that mind had but one variable—time. Without accepting his conclusion, we can concur in the importance of the time factor in mental happenings, and appreciate the part played in psychology by the Wheatstone-Hipp chronoscope, stop watches and other instruments for measuring and recording time intervals. Starting from comparatively simple beginnings, reaction times of more and more complex mental activities were determined, and the chronoscope has become for the psychologist what the balance is for the chemist.

It would seem undeniable that intensity is a characteristic of sensory experience. When we attempt to evaluate this quantitative element in sensation we are met with difficulties. First comes the fact that the same material object may be perceived by sensations that vary greatly in intensity. The intensity of experienced sound of an express train is as a rule different for a passenger comfortably sitting in it and for a would-be passenger standing on the platform as it rushes through a station. The intensity of the noise of a ticking clock sinks below the threshold when we are engaged in an absorbing task, but the energy output of the clock remains reasonably constant. The same source of stimulation occasions different intensities of sensation according to the subjective attitude of the recipient of the energy falling upon the sensory organs. To secure consistent results it is necessary to control the conditions, and in experiments involving sensation-intensity this is done by arranging that the subject shall give undivided attention to the stimuli. The maximal intensity thus obtained is accepted as constant for the given stimulus.

Normally we do not discriminate the intensity of the sensation, nor are we interested in the particular sensation itself. Our interest may be in the pleasure tone of the whole sensation complex, as when enjoying music or a glorious bank of flowers, but, more often, our interest is in the properties of the external objects that occasion the experience. The lightning's flash, the trumpet's

blare are ominous portents, rather than intense sensations. This preoccupation with the things of the external world has given us science but makes introspective analysis of experience difficult. It endows sound, light and matter with 'intensity' which we proceed to measure with acoumeters, photometers and a balance. For the psychologist, however, these measurements are only of interest so far as they may help to throw light on the problems of sensory experience.

Early in the life of experimental psychology the smallest change in the stimulus value that could be appreciated in sensation became the subject of inquiry and, for the determination of this difference threshold, difference limen, least perceptible difference or just noticeable difference, the psycho-physical methods were developed and gave to psychology a technique the wide serviceability of which does not even yet seem to be fully appreciated.

[Mr. Bartlett here discussed the Weber ratio.]

PHYSIOLOGICAL INVESTIGATION

Another branch of experimental psychology busies itself with the physiological concomitants of emotional experience. On a smoked drum record of a subject's breathing, measurement can be made of frequency and amplitude, and changes therein can be compared with introspection records to find out their meaning. The inspiration-expiration ratio, with varying emotional states, shows marked variations from the average value, rising to as much as three times that value in 'amazement' and sinking to about half the value under the sense of guilt or shame awakened by conscious misrepresentation or wilful falsehood. These changes in the ratio of inspiration to expiration are accompanied by marked changes in frequency, amplitude and form, which result in characteristic curves that the eye can read and interpret without recourse to the millimetre scale and time record. The whole profile, rather than its analysis, is diagnostic.

INTELLIGENCE TESTS

An entirely different method of measurement had by 1920 reached the stage where it could not be ignored. Intelligence tests were being widely used, and the Board of Education set up a consultative committee to report "on psychological tests of educable capacity and their possible use in the public system of education". The report appeared in 1924, and the principal findings of the committee were: "That up to the present the only kinds of psychological tests of educable capacity that have been sufficiently developed to

be of much service in schools for the purpose of diagnosing and assessing such capacity are tests of 'intelligence', standardized scholastic tests, and to a less extent vocational tests". Let us consider the first of these.

The foundations of the method are laid in Binet's happy thought of 'mental age' and Spearman's application of methods of correlation to psychological data. Prior to Binet's work, testing had been confined almost entirely to tests of sensory discrimination. Spearman proceeded to calculate correlation coefficients between the results of such tests and estimates of intellectual ability. Binet, "instead of measuring the intensity of simple faculties", set out by tests "manifold and heterogeneous . . . [to] measure acts of adaptation . . . to determine how many years an individual is advanced or retarded". His measure is a time measure based on the normal development of intellectual ability. Originally his "manifold and heterogeneous" tests were 54 in number, used for the diagnosis of mental deficiency. This number has steadily increased until, in the New Revision of the Stanford-Binet Scale, issued by Terman and Merrill in 1937, there are two sets of 129 tests each, covering mental ages from 2 years to 22 years 10 months and providing for Intelligence Quotient values up to 170 for children and up to 152 for adults.

By correlation methods, Spearman compared the results of tests of sensory discrimination and estimates of intellectual ability, inaugurating a new phase of mathematical activity in psychology, in which, again, the tasks led to new methods. The correlation between two sets of measurements or rankings is a measure of the extent to which the sets of figures agree, and this agreement would seem to be due to a common factor in the abilities measured. It follows that "when any pair of abilities are to any extent correlated with each other, to this extent *they can be regarded as depending upon a common factor*. . . . Otherwise expressed, each of the abilities may be taken to involve two factors, the one common to both, the other specific to that ability alone". Extending this thought to more than two abilities, Spearman developed his theory of two factors, and evolved various criteria for establishing the existence of the relations that fit the theory, culminating with the tetrad equation, $r_{ap}r_{bq} - r_{bp}r_{aq} = 0$, and its probable error.

VOCATIONAL TESTS

The Educable Capacity Report of 1924 gave limited approval to vocational tests, but concluded that "the range of such tests [was] not yet sufficient for any recommendation". The Medical Research Council, through the Industrial Fatigue Research

Board, had been busy with the problem for some years, and in 1922 this Board, in co-operation with the National Institute of Industrial Psychology, undertook a preliminary investigation into the possibilities of vocational guidance. The group of psychologists who, under the leadership of Prof. Cyril Burt, carried out the investigation, reported in 1926. The research had shown that both stability of employment and satisfaction are greater among those who had obtained employment in keeping with the recommendations made on the basis of the test results than amongst those in other employment, and that it would seem true that the methods used "will prove of the utmost value to the individual and to the community, to the employer and to industry as a whole".

Since 1926 much work has been done on vocational guidance, both by the Board, under its new title of Industrial Health Research Board, and by the National Institute of Industrial Psychology. As a result of investigations carried out by these bodies and many individual research workers, in Great Britain and other countries, tests for abilities and skills of value in industry, and for particular disabilities that unfit for certain occupations, are available and are being used with increasing frequency.

Much progress has also been made with the measurement of temperament and character, a valuable summary of which will be found in Report No. 83 of the Board. In this report, Vernon reaches the conclusion that, while tests to measure attitudes or interests have not the objectivity and accuracy achieved by tests for abilities, and character and temperament traits present even greater difficulties, good progress has been made, and, with a better understanding of the methods and meaning of the results of factor analysis, "we may hope before long to achieve a fairly complete classification of all our psychological measuring instruments which would be of the utmost value in many branches of pure and of applied psychology".

The order of progress has been: the measurement of intelligence; the measurement of special abilities and skills; the measurement of attitudes and interests; the measurement of temperament and character traits. To deal with the growing complexity of the material, mathematical theory and technique have been developed until sometimes there seems danger of losing the facts in the figures, and it becomes necessary to remind ourselves that mathematical analysis cannot confer value on valueless data. The methods by which the data are collected are of prime importance. In the earliest work the instruments used were the ordinary scales of physical science; what a

man did or judged was submitted to measurement in millimetres, grams, foot-candles, pitch, thousandths of a second, ohms. Physical measurement was possible, and, too often, was equated to the mental experience.

When physical measurement is impossible, we may be able to secure rankings. Most people can arrange colours, pictures, persons, poems or holiday resorts in an order of preference.

[Mr. Bartlett then discussed ranking, grading and rating.]

PERSONALITY CHARACTERISTICS

We pass from attitudes towards persons and things to the question of personality characteristics. Here, as a rule, the scale that we use is completely subjective, the judges' opinions of the extent to which the characteristic is present being expressed in some recordable form. The judge may decide that his own or another's 'adaptability' is poor, fair, good, very good or excellent, and these classes may be turned into numbers 1 to 5. He may class a person as schizophrenic, schizoid, schizothyme, neutral, cyclothyme, cycloid, cyclophrene. He may rank persons by the usual techniques in order of merit or assess value by giving marks up to 10 or 20. He may approximate to an objective standard by the so-called 'man to man' scales, in which two persons are selected as exhibiting the trait in extremely high and extremely low form, respectively. Then a person is chosen that falls halfway between the extremes, and then two others halfway between the central person and the extremes. We have thus a scale of five reference values, secured by a method similar to the psychophysical method of equal appearing intervals. In using such a scale we shall find that many of those we wish to judge fall between two of the scale values and, as with a transit instrument or other reading, we find ourselves estimating halves and quarters and, with increased practice, tenths. Or again, we may objectify the distance between the scale items as lines and mark the point on the line that corresponds to the position taken by the person in the mental scale.

When a number of characteristics have been rated, they may be brought together graphically in the form of a profile, and such profiles have been found of great value in vocational selection and guidance. Instead, however, of measuring a number of individual characteristics and then bringing them together in a profile, we may rank or grade the various characteristics among themselves for a particular person, and compare that person with another by calculating the correlation between the rankings or grading of their respective characteristics. Further, if we assume normal

scatter and throw each individual set of traits into identically sized groups balanced about the central value, we can calculate the correlation by the very simple formula,

$$r = 1 - \frac{\Sigma (X - Y)^2}{K},$$

where X and Y are the actual grade values given to the various items and K is a constant equal to half the possible maximum value of $\Sigma (X - Y)^2$. The method of correlating persons had been used for some years without its essential difference from correlation of tests being clearly stated. It is to the credit of Stephenson that in his papers on "The Inverted Factor Technique" he has defined these differences and directed attention to the advantages of the method.

We have sampled the methods and results of experimental psychology and trust that the sample is a fair one, though possibly somewhat overloaded with ancient history that has become of interest to physicists in contact with the human factor. The sample is statistically inadequate, but has, we hope, enabled the mind to perform its accustomed trick of reaching conclusions from cues or

data logically and mathematically altogether insufficient. We have seen psychologists using physical measurements and developing methods of dealing with the new forms of data so obtained. We have seen them following intuitive perception in the belief that their material had quantitative elements not at present amenable to the measurements of the physicist, replacing the scales of physics by estimates, ranking, grading and rating scales and developing mathematics to deal with these new data. We have seen them progressing step by step into more and more difficult fields of inquiry, and may look forward with confidence to yet greater achievements. In these circumstances, does it matter much if some continue to believe, with Malebranche, Leibniz and Kant, that our data contain nothing "that can properly be called measurements" and that it is presumption for us to think that, in any reasonable sense, our data, theories, methods and results constitute "a systematic science"? After all, there is a sense in which logical and mathematical proofs are what the psychology of advertising has called "rationalization copy". Scientific insight, as everyday perception, has ever run ahead of measurement and mathematical proof.

OBITUARIES

Prof. Edward Westermarck

EDWARD WESTERMARCK, whose death at the age of seventy-seven is reported from Finland, must always rank as one of the world's great anthropologists; and, moreover, in view of the fact that his first and perhaps most famous work "The History of Human Marriage" appeared so early as 1891, only one year later than "The Golden Bough", he can be counted among the actual pioneers of anthropology in its social aspect. Born in 1862 at Helsingfors, he might be said to be connected since his birth with its University, his father being the bursar and his mother a daughter of the librarian; and here, having graduated and taken his doctorate with a thesis on primitive marriage, he became lecturer and afterwards professor of philosophy; only deserting his *alma mater* after the War of 1914-18, when he was appointed by a liberated Finland to organize as Rector a new Swedish-speaking university at Abo.

Meanwhile, Helsingfors had always treated Westermarck with great liberality, providing him with travelling scholarships in five different years so that he might study, first in England, and later in Morocco; and afterwards tolerating an arrangement whereby, when he had served as professor in Finland during the winter months, he could spend the summer term in London as lecturer at the School of Economics and afterwards as holder of a Martin White chair of

sociology. Term over in London, he would slip off quietly to Morocco, where he was as much at home as at Box Hill; and for his friends it amounted almost to a game of thimble-rig to guess at any moment where his elusive presence might be discovered.

Westermarck's first book, a masterpiece of detailed research, proved him not only to be exceedingly well informed—he had the reading-room of the British Museum to thank for that—but likewise to be a trained philosopher whose empirical leanings had led him to pay special attention to the theory of evolution. Indeed, he tells us in his autobiography ("Memories of my Life", 1929) that already by 1887 he had decided to write about human marriage on lines suggested by Darwin's "Descent of Man", and with special reference to the problem "how to explain the veil of modesty and discretion that is thrown over the sexual life". Further, Alfred Russel Wallace helped him in the preparation of his argument with critical advice, and went on to furnish his book with an interesting preface in which he praises both its thoroughness and its philosophic quality. This subject, then, Westermarck made his own and his second and much enlarged edition in three volumes (1921), as well as the "Short History of Marriage" (1926), must always retain high authority.

A second *magnum opus*, "The Origin and Development of the Moral Ideas", is also a classic, even if