THE PSYCHOLOGICAL BULLETIN

GENERAL REVIEWS AND SUMMARIES

INTROSPECTION AND GENERAL METHODS

BY KARL M. DALLENBACH

Cornell University

What is introspection? Pepper (9) in his article of that title takes up the answer systematically. He begins his paper by considering the influences which have tended to differentiate introspection from the method of the objective sciences; and then takes up the methods historically showing; first, that introspection is the older, indeed, "That the objective method developed out of introspection;" and secondly, that there is no qualitative difference between them. The latter is his thesis, which he develops (1) by the exposition and criticism of Titchener's analysis of introspection, which he, in opposition to Titchener, shows to be equally applicable to the objective methods; and (2) by the consideration of the kinds of processes studied, which furnishes him with a cue for the differentiation, which "is simply this, the objective data are picked, whereas introspective data are anything that comes along." Following this lead, he finds that in "a thoroughly developed objective science nothing is acceptable but visual data," whereas "the introspective method will accept any kind of experiences whatsoever as fit material for its processes." His conclusion is: "the distinction between the objective and introspective methods resides in the kind of processes described."

Of his thesis, that there is no qualitative difference between the two methods, there is no criticism; but not so of the arguments advanced in favor of his conclusion. They show a complete misunderstanding of the subject. His criticism of Titchener reveals a like condition; it is valid only in so far as it rests upon his exposition of Titchener which is garbled beyond recognition.

After raising and meeting the objections against introspection brought forward by certain behaviorists, Laird (7) arrives at the conclusion that "it seems both legitimate and necessary to assume that introspection has the same general characteristics as any other mental process by means of which we are able to apprehend the truth of fact." He then comes to the thesis of his paper which is that introspection, which is a "kind of cognition," ought to be regarded as a legitimate method of apprehending the mental phenomena "in spite of the numerous objections to this view which are either expressed or implied in contemporary psychology and philosophy."

The objections usually raised against introspection are classified and discussed under three heads: (1) its possibility, (2) its fallibility, and (3) its necessity. Laird's conclusions regarding these questions are: (1) that it is possible; that "there is no intrinsic absurdity in supposing that introspection is a process of direct cognitive acquaintance with our own minds;" (2) that it is not necessarily fallible; "Nothing can be transformed in any sense whatever simply owing to the fact that it is known," and (3) that it is necessary if we are to know the mind for "... introspection is the only means of direct acquaintance with the mind," because none of "... the characteristic of consciousness can be discovered by other than introspective methods."

After giving a brief history of psychology, Lalande (8) considers next the province, the objects, and the methods. The province of psychology is discussed under four headings: "(1) *Psychologie de réaction*; (2) *Psychologie de conscience ou de sympathie*; (3) *Psychologie réflexive et critique*; and (4) *Psychologie ontologique*."

Under methods, he takes up the introspective, the comparative, and that of the laboratory. The objections to introspection, he finds, "sont beaucoup plus théoriques que pratiques," and "elles ne détruisent pas la possibilité de l'introspection." In order to insure greater accuracy in introspection Lalande lays down five precautions: (1) the observed facts should be recorded as soon as possible; (2) many observations should be made in order that general results may be distinguished from individual; (3) the questionnaire and interrogatory methods do not give the psychological representations or the reasons which produced the subjects' responses, but simply crude results; (4) first-hand knowledge of the phenomena described is very useful and perhaps indispensable for a correct interpretation of the results; and (5) since auto-suggestion,

even with the most critical minds, is extremely easy, the reports should be surrounded by such conditions that the observers are unaware of the expected results. A separate chapter is devoted to the introspection employed in experimental work.

Lalande next treats the comparative methods, which he discusses under the following headings: (1) the pathological and psycho-analytical method; (2) the sociological and genetic methods; and (3) the methods of animal psychology. After which the most common of the laboratory and statistical methods are mentioned, and the article is concluded by a discussion of terminology.

Bourdon, in a short article (4), has attempted to reopen the controversy which for many years centered about Fechner's formulation of Weber's law. After pointing out the well known fact that Fechner's formulation differed from Weber's, he turns to his thesis, which is to show that the distinction and opposition of sensation and stimulus is an error. Both are alike; they are one and the same thing. He points in proof of this to the fact that "*Le psychologue et le physicien opèrent . . . incontestablement, sur les mêmes phénomènes et procèdent essentiellement de la même manière.*" In order to establish this argument, he reviews the traditional experiments—the shadow experiments and the lifted weight experiments—and shows that the psychologist in each and every one of them is judging the same phenomena and judging them essentially in the same way as the physicist. His conclusions therefore are that "*La loi de Weber ne doit pas être formulée de telle manière qu'elle implique une opposition de la sensation et de l'excitant;*" that "*La loi de Fechner doit être rejetée,*" for "*Fechner a mesuré, sous le nom de sensations et d'excitants, les mêmes phénomènes;*" and that "*Les sensations, en tant qu'opposées aux excitants, sont des fictions.*"

This article might well have appeared during the heat of the historical controversy, but it is difficult to understand why any one conversant with the psychophysics of the present day should take such a position.

The object of Thomson's (12) paper is "to show how the Constant or Fechner-Müller-Urban process of calculating thresholds can be deduced from first principles." He, in his opening paragraph, summarizes the article as follows: "The historical development of the process is first traced, and then contrasted with direct deduction. The latter throws into clearer relief the nature of the assumptions and approximations made, and justifies Urban's, as

against Müller's, Table of Weights." The article is apparently an excerpt from the paper with which Thomson won the prize of \$100 offered in 1914 and awarded in 1919 for the best paper on the Availability of Pearson's Formulæ for Psychophysics.

Johnson (6) publishes a note on Dunlap's method of determining the mean variation, in which he calls attention to a misprint in the original article of a *plus* for a *minus* sign in one of the formulæ; points out that not only the application but also the method itself was developed by Dunlap; and adds another rule to the two given by Dunlap.

Boring has contributed three articles (1, 2, 3) to the general subject of statistical treatment of data. The first of these (1), "Mathematical *vs.* scientific significance," deals with the "apparent inconsistency between scientific intuition and mathematical result." After briefly reviewing the customary methods of determining the "significance" and "probability of difference," he mentions the common experience of those working with human data that "these methods frequently give values that are, on the basis of professional opinion, too high." This particular attitude is, he maintains, not due to the unreliability of professional opinion, but to the fact that the samples from which the "significance" is computed, even with the best methods of selecting the "unselected data," are frequently not representative of the total group. So the competent man, cognizant of this, "very often discards his mathematical results because in his judgment the mathematically 'significant' difference is nevertheless not large compared with what he believes is the discrepancy between his samples and the larger groups which they represent." All of which merely emphasizes the point with which Boring is concerned, namely that "statistical ability divorced from scientific intimacy leads nowhere."

In his second article (2) on "The logic of the normal law of error in mental measurement," which he tells us was the outgrowth of certain perplexing questions that arose in the Surgeon General's Office U. S. A. in connection with the ratings of recruits in the Army intelligence tests, Boring examines, first, the *Nature of probability*, considering it both historically and logically; and then the *Rôle of the Law of Error*, which he regards as "both an inspiration and a limitation in statistical measurement." This, too, he treats historically, considering its formulation by Laplace, its development by Gauss, its uncritical extension by Quetelet, its *a priori* application by Galton and the English school of biometricians,

the reaction against it lead by Lexis and Westergaard in Germany and to a lesser degree by Galton and Pearson in England; and he thus comes down to the present, where he considers its naïve and oftentimes unwarranted acceptance. A rather flagrant case is cited, and criticized in his third article (3) which is on the "A priori use of the Gaussian Law."

It is against the blind acceptance of the normal law, the Gaussian fetish, that he inveighs, and the sole purpose of this part of the second article is to remind the reader that there is no "magic to the normal law," and that there is no a priori reason for expecting it, but that, quite on the contrary, "we have considerable reason to expect not to find it, since the form of distribution depends on the unit of measurement, and we have arbitrarily chosen one of a possible infinite number of units."

After mentioning in illustration two experiments in which this problem arose,—in one the limitation of the normal curve was recognized and the observed facts were adhered to, and in the other there was a naïve acceptance of the law and a calibration of the material affected by reference to it,—he proceeds to show that belief in the universality of the normal law is due to analogy, which to a certain extent is justifiable in (1) games of chance and (2) digital expansion of numbers, where within each universe the same scale of units is throughout employed. In the distribution of (3) physical measurements, (4) mental measurements, and (5) psychophysical judgments the predication of the normal law is entirely another matter. With (3) the physical measurements the generalization is 'approximately' correct, "for a certain amount of alteration of the scale still leaves a sensibly normal curve;" but in (4) mental measurement, where a unit is still wanting, "we are very far . . . from a general conclusion that intelligence is normally distributed, and still farther . . . from stating that mental abilities follow the normal law." With (5) psychophysical judgments the generalization is permissible, when the scale and the judgments are related, as in the cases when the reports are judgments of stimulus, but not in "introspective" psychophysics where the judgments are given in terms of sensation, and are thus in no way related to the physical scale of units chosen.

For work in mental measurement there are four logical possibilities: (1) The a priori determination of the form of distribution,—a popular course, but without sanction and to be resisted; (2) the determination of the distribution from the psychological unit,—but

no applicable psychological unit has as yet been found; (3) in lieu of this, the determination of the distribution by some arbitrary "physical unit" such as year (mental age), second (time-measure), item (task test), gram (lifted weights), etc.,— but a physical unit is not a mental unit and can not be applied psychologically; and (4) the determination of the rank-orders of our psychological quantities,—the only procedure which at present is justified, since no unit is presupposed, and though the results are less intricate than those obtained by any of the other alternatives, a picture is represented by the serial constants "which, though rough, truly represents the rough material which they describe."

This series of papers can be read with profit by every student of statistical methods, particularly those working in the applied fields, in the standarization of mental tests, etc., for it is they who most frequently succumb to the "Gaussian fetish."

Another appeal to first principles in mental tests is made by Ruml (11), in an excellent article on "The Need for an Examination of Certain Hypotheses in Mental Tests." Deploring the fact that little of theoretical value has come from the work in mental tests, he calls into question three of the fundamental hypotheses which serve as the basis for all testing work.

He questions, first, the belief that general intelligence, even in quantitative terms, can be expressed as a linear or one-dimensional function; secondly, the supposition of linear regressions between test performance and general intelligence; and thirdly, the assumption of a static intelligence level, an assumption which "is necessary to mental test work as it is now conceived."

His explanation of the failure of theory to keep abreast of practice is: (1) That it is due to the historical derivation of mental tests; (2) that it is due to the inertia of thinking; (3) that it is due to bias, which may be attributed "to the nature of the terms in which the results have been expressed," and to the type of analysis "which our limited and frequently misused statistical technique makes possible"; and (4) that it is due to a willingness "to accept statistical hypotheses as applied to intelligence simply to have statistical technique available for use."

Herring (5), in an article which is a continuation of an earlier work, explains the method and technique that he employed in the development of his scale to measure abilities in scientific thinking.

The paper is divided into six parts. In Part I. he sets forth the principle under which he works, "that mental abilities are

distributed as are physical traits," *i.e.*, they are distributed according to the probability curve, and the assumptions made under it, which are, "first, the distribution of the abilities to be studied is normal; and second, the variability of each large randomly selected group of pupils is approximately the same as that of any other." In Part II, in order to establish the validity of his method, he considers the probability curve, explains the P. E. and his use of it as a unit of measure; and defines the terms and relations which he uses. In Parts III, IV, V, and VI the steps employed in the derivation of the scale are completely described.

Herring's exposition is very clear and definite. It is regrettable, however, that he did not spend more thought upon fundamental considerations. As it is, the article is a good example of the kind of work criticized by Boring and Ruml. The hypotheses upon which he constructs his scale are naïvely accepted as if incontrovertibly established.

Thurstone, in the exposition of his examination for College entrance (14), discusses the methods of evaluating the tests. His plea is for practicality, and to this end he discards the method of multiple correlation as a means of analysis and employs in its stead the method of "critical scores." This method consists "simply in plotting a scatter diagram for each test, showing relation between scholastic attainment and mental test score for each individual. An upper critical score is selected such that all who score above it are above the average in scholastic attainment. A lower critical score is selected such that all who score below it are below the average in scholastic attainment. . . . When the tests are combined . . . the median percentile rank is corrected by adding five points for each test in which the subject is above the upper critical score and deducting five points for each test in which the subject is below the lower critical score." The question that automatically rises at this point is: why five? Thurston offers no explanation for his selection of this value. It seems, however, to be satisfactory. The chief advantages of the method are: (1) it is more controllable than the method of multiple correlation for establishing diagnostically significant criteria; (2) it is easily applied; and (3) it emphasizes "prognosis with respect to each individual student rather than the abstract statement of relationship between the variables concerned."

Thurstone's plea touched a sympathetic chord in Pressey, who in a note of appreciation (10) offers three suggestions which would

further "the consideration of the practical problems of individual diagnosis in evaluating mental tests." The first deals with the method of evaluating the tests, and the second and third with the form in which the findings of the examination should be reported.

Another article which questions the tenets of technological psychology is published by Thorndike (13). It is on the constant error in psychological ratings.

Thorndike finds from a review of number of studies on ratings—studies made of employes in two large industrial corporations, of cadets in the aviation corps, of officers in the army, and of school teachers—that there is a "constant error toward suffusing ratings of special features with a halo belonging to the individual." He arrives at this judgment because the correlations between the different traits estimated "are too high and too even." "Obviously a halo of general merit is extended to influence the rating for the special ability, of vice versa."

As a consequence of this study Thorndike "has become convinced that even a very capable foreman, employer, teacher, or department head is unable to treat an individual as a compound of separate qualities, and to assign a magnitude to each of these in independence of the others." Therefore, "science seems to demand that, in all work on ratings for qualities, the observer should report the evidence, not a rating, and the rating should be given on the evidence to each quality separately without knowledge of the evidence concerning any other quality in the same individual."

REFERENCES

1. BORING, E. G. Mathematical vs. Scientific Significance. *PSYCH. BULL.*, 16, 1919, 335-338.
2. BORING, E. G. The Logic of the Normal Law of Error in Mental Measurement, *Amer. J. of Psychol.*, 31, 1920, 1-33.
3. BORING, E. G. A Priori Use of the Gaussian Law. *Science*, N.S., 52, 1920, 129-130.
4. BOURDON, B. La loi de Weber et celle de Fechner. *Rev. phil.*, 88, 1919, 119-121.
5. HERRING, J. P. Derivation of a Scale to Measure Abilities in Scientific Thinking. *J. of Educ. Psychol.*, 10, 1919, 417-432.
6. JOHNSON, B. Dunlap's Method for the Mean Variation. *Psychobiol.*, 1, 1918, 325-326.
7. LAIRD, J. Introspection. *Mind*, N.S., 28, 1919, 385-406.
8. LALANDE, Q. La psychologie, ses divers objets et ses methodes. *Rev. Phil.*, 87, 1919, 177-221.
9. PEPPER, S. C. What is Introspection? *Amer. J. of Psychol.*, 29, 1918, 208-213.
10. PRESSEY, S. L. Suggestions with Regard to Professor Thurstone's "Method of Critical Scores." *J. of Educ. Psychol.*, 10, 1919, 517-520.

11. RUMML, B. The Need for an Examination of Certain Hypotheses in Mental Tests. *J. of Philos., Psychol., &c.*, 17, 1920, 57-61.
12. THOMSON, G. H. A Direct Deduction of the Constant Process Used in the Method of Right and Wrong Cases. *Psychol. Rev.*, 26, 1919, 454-464.
13. THORNDIKE, E. L. A Constant Error in Psychological Ratings. *J. of Appl. Psychol.*, 4, 1920, 25-29.
14. THURSTONE, L. L. Tests for College Freshmen. *J. of Educ. Psychol.*, 10, 1919, 129-142.