THE PSYCHOLOGICAL REVIEW.

THE CLASSIFICATION OF PSYCHO-PHYSIC METHODS.

By DR. EDWIN B. HOLT.

Every one who has undertaken experiments on the relation of stimulus to sensation, or even read considerably on the subject, must have become aware of the inadequacy of the historical and still current classification and designation of the psycho-physic methods. For many cases arise in practice which have no immediate place in the classification, as for instance those in which it is necessary to recognize and more especially to evaluate judgments other than those of 'less,' 'equal' and 'greater' (as say 'much less' 'uncertain' and 'much greater'); and other cases arise whose place is doubtful because they have features of several of the methods but have not all the essentials of any one of them.

If for instance one were to wish to find the acuteness of a given sense in an interesting pathological patient, and if from independent reasons (as may well happen) one were obliged to adopt the so-called method of right and wrong cases, one might be brought to pause, if the subject were impatient or perhaps hysterical, by the appearance of adverse emotions and fatigue, due to the almost endless repetition of just two stimuli which this method requires. And yet perhaps one could get from the subject one fifth the required number of judgments on each of five different pairs of stimuli, or else one tenth of the number from ten pairs. Now clearly if one were careful to have one stimulus common to these five or ten pairs, one could calculate by the method of least squares from the five or ten resulting

groups of judgments the coefficient of precision ('*Präcisionsmass*') with as much accuracy as otherwise from one group of ten times as many judgments, on a single pair of stimuli. Yet in such an instance one would not be using the method of right and wrong cases. For although the point of departure and the answer yielded conform to the descriptions of this method, yet the actual data gotten and used in calculating the result are precisely such as one would have gotten from the method of least differences, or minimal changes (*Abstufungsmethode der kleinsten Unterschiede*) if peculiarities of the sense-organ investigated or exigencies of apparatus available had forced one to use few and coarse gradations of stimuli and to present the different

FIG. 1.

pairs at random. For in both cases the data would consist in judgments of 'less' (<), 'equal' (=) and 'greater' (>), which would be distributed as shown in Fig. 1, where the abscissæ represent the value of the stimuli greater or less than the standard stimulus (0) which is common to the ten pairs, and where the ordinates represent the number of judgments.

In short the experiment described does not come under any one of the traditional methods, since it has also features of at least one other method, nor can it properly be called a 'combination' of the methods, since it has not all the essential features of any one.

Apart from the confusion which every reader or experimenter in psychology has probably felt in assigning a given experiment to one of the four or six 'methods,' the unsatisfactory state of the whole matter is amply witnessed by the historic errors made by the most distinguished experimenters in designating the methods they have used, by the confused and overlapping definitions of the several methods invariably given in text-books, and by the amendments offered from time to time by various theorists. The errors of designation, although several were made by Fechner himself¹ are chiefly of historic interest. Of systematic interest, however, are the current definitions and descriptions of the psycho-physic methods, and the several emendations which their inconsistencies have elicited.

The case above cited of an experiment which, by having features of two methods and yet lacking the essentials of any one, belongs properly under none of the four or six methods, is not wholly unprovided for in the treatises. Thus Wundt,² in defining the method of mean gradations, describes, for the case that the middle stimulus varies irregularly, a certain simple way of deriving the mean value from the raw data: but this way need not be used, he says, since the raw data can also be treated by the method of right and wrong cases, in which case the whole procedure becomes 'a combination of the method of mean gradations with the method of right and wrong cases.' Now the simpler way of evaluating the data applies exclusively to the method of mean gradations: is it then essential to this method? Clearly not, since it may be abandoned in favor of the method of right and wrong cases. But then the method of mean gradations is not, as is generally supposed, a method for both getting and then evaluating raw data, but only for getting them. Conversely, too, the method of right and wrong cases is not at all a method for getting data but only for treating them when gotten. But if this is true, these two methods are not alternatives to be chosen between, but supplementaries to be used in combination; a conclusion which is at variance with the present theory and practice.

¹ Müller, G. E.; *Zur Grundlegung der Psychophysik*, Berlin, 1879, S. 56 ff.

² Wundt, W.; *Grundz. d. physiol. Psych.*, 5te Aufl., Leipzig, 1902, I Bd., S. 480.

This contradiction arises, of course, from the confused definitions of the methods, whereby these latter, instead of being four mutually-exclusive logical classes, as they should be, are defined as being now exclusive, now partially identical, now coördinate, and now not coördinate. And this confusion extends through the whole tissue of the methodology, at least the Wundtian. For this author, after dividing all methods in two (mutually-exclusive ?) groups, of gradation (*Abstufung*) and of telling-off (*Abzählung*), explains¹ that 'Among the telling-off methods the method of mean errors is most nearly related in its origin to the gradation methods or more particularly to the method of minimal changes.' Indeed,² 'The method of mean errors arises from the method of minimal changes in case one limits oneself to taking the just not-perceptible differences of stimuli.' By this is presumably meant that the raw data in the method of mean errors are identical with a part of the data in the method of mean gradations; and the fact is left out of account that the data are further quite differently treated in the two cases, in ways which by no means 'arise' the one from the other. In short Wundt's systematic treatment of psycho-physic method stops short of logically exact and consistent definition.

Yet these ambiguities are by no means peculiar to Wundt: firstly because most other modern treatises follow, or perhaps even copy, the veteran psychologist, secondly because the same ambiguities have prevailed from the very first. Thus while Wundt finds the method of minimal changes merging into that of mean errors, Fechner³ found that it 'goes over' into the method of right and wrong cases! The key to the situation is this: the four psycho-physic methods are historical developments bearing the marks of their growth and of the accidents they have met. The methods have not been defined, they have been *used*; and where a user has had to modify a method he has generally not modified its name, so that there are several varieties under every method, and these bear the most diverse and unsystematic relations to one another. So great has been

¹ *Ibid.*, S. 473.

² *Ibid.*, S. 472.

³ Fechner, G. T.: *Elemente d. Psychophysik*, Leipzig, 1860, I Th., S. 75.

the dignity of tradition in this matter of method, that a thorough-going revision and consistent systematization of the procedures has never been achieved.

To systematize the procedures is not necessarily to analyze or revise their mathematical details,—an ambitious performance indeed,—but the frame-work of methodology can be reconstructed in and for itself. This frame-work so recast will carry with it undisturbed, as tent-poles the canvas, the vast multitude of details.

To commence, one must first survey the methods in their confusion, then either discover or postulate some one feature as the sole essential of each method, and then deduce the logical consequences. The Wundtian account is the most characteristic and may well be made the starting point.

The methods fall in two groups, those of gradation and those of telling-off. The former group has two classes, the method of minimal changes and that of mean gradations; the latter also two, the method of mean errors and that of right and wrong cases.

I. THE METHOD OF MINIMAL CHANGES.¹

“In this method one seeks to determine at different points on the scale of stimuli such a change in the intensity of stimulus as produces a barely perceptible change of sensation.” And, in detail, one finds what interval between the standard stimulus and a lower, compared stimulus is needed to make the two feel different, and the same between the standard and a higher stimulus; and again the interval is found between standard and lower or higher stimulus which will just not make the two feel different. The average of these four intervals is the threshold wanted.

This is good as a scheme, but on coming to the practice one finds that there is no interval which just is or is not *always* felt as a difference. There are intervals which are generally felt as one or the other, but in order always to be so felt the interval must be either so large or so small that it tells nothing about the accuracy of discrimination which one is trying to find. The scheme appears to be applicable if the compared stimulus can

¹ Wundt, *op. cit.*, SS. 470, 476-479.

vary about the standard *continuously*, for then one commences with the compared stimulus plainly larger, say, than the standard and diminishes the former until it no longer seems larger; one calls this interval the upper threshold of difference, and finds the other three desired intervals similarly. But the continuously applied stimulus tires the sense, while habit retards or expectation hastens (presumably according to the temperament of the subject) the moment in which the relation of the stimuli is felt to change; and it has not been proved that these factors cancel themselves out. Thus the result yielded by such a procedure would be modified by the temperament of the subject and the capacity of his sense-organ to resist exhaustion, and would not be a pure measure of his discrimination.

If the compared stimulus is such as cannot be varied continuously then the procedure in question is, for all careful work, out of the question, for the final result will be found to be not a little dependent on the size of the step-wise gradations which one arbitrarily has adopted in the series of compared stimuli. And furthermore the case will often come up and have somehow to be taken into consideration that one interval will yield a perceptible difference, the next smaller will not, while the next smaller than that will again do so.

In short, whether the compared stimulus is to be varied continuously or step-wise, it is advisable (as Wundt himself admits; S. 478) not to present these stimuli in their orderly progression, but rather in a random succession. But 'this procedure has at the same time the character of a combined method, since it approaches the telling-off methods.' Whether 'combined' or not, it yields three kinds of judgments ('greater' 'equal' and 'less'), and they show the arrangement which we have met before in Fig. 1.

What is now essential to this method of minimal changes? It is not the use of a standard stimulus and compared stimuli, for the method of mean error also uses these; nor is it the 'minimal' gradations in the compared stimulus, for the methods of mean gradations and of mean error require these; nor is it yet the ascertainment of the intervals above and below the standard stimulus which barely do *not* give a difference in sensation, for

the method of mean error involves also this. Essential to the 'method' of minimal changes is only the *project* of finding that difference between stimuli 'which produces a barely perceptible change of sensation.' This, however, is not a method, but a problem: furthermore the quantity so sought needs a precise definition (which the phrase quoted by no means gives) and gets this only from the method finally fixed on for solving the problem. Now the methods of evaluating data described by Wundt as peculiar to the 'method' of mean gradations have been shown to be inadequate for careful work; and in order to be precise, as Wundt himself advises, one must use a method which 'approaches' (to put it plainly *is*) the method of mean error. In a word, the 'method of minimal changes,' is not a method at all, but a problem which one sets oneself. It can be solved with precision only by such a procedure, and by any such, as yields at least three classes of judgments grouped as in Fig. 1. How from these a solution is to be gotten is not told by the 'method of minimal gradations.'

As if so much confusion were not enough, the account provides only for the case of finding that change of stimulus which produces a change of sensation, that is, the threshold of *discrimination*; whereas the case that one wishes to get the threshold of *sensation* may properly come under one of the methods, and certainly can come under no other than this of minimal changes. Finally this threshold of discrimination is an average or two thresholds, those of just perceptible and just imperceptible difference, each of which has often been made the basis of investigation, although in either case the method has generally been called minimal changes. Thus the 'method of minimal changes' not only is a problem and not a method; it is not even a definite, single problem.

2. THE METHOD OF MEAN GRADATIONS.¹

This method, which Wundt substitutes for the older, somewhat more natural, and at least equally important 'method of over perceptible differences,' consists in finding a mean stimulus which shall seem to lie equally far from two chosen extremes.

¹ Wundt, *op. cit.*, SS. 471, 479-481.

Two intervals are to be made equal. 'But in order to get reliable results,' Wundt says, 'it is necessary to combine this method either with that of minimal changes or with one of the two telling-off methods about to be described.' In fact it appears that when the mean, variable stimulus is presented with the two standards, it will seem to lie nearer sometimes the lower, sometimes the higher, standard; sometimes just half-way between. This gives rise to three groups of values for the mean stimulus, those for which it is judged to lie more and less than half-way and just half-way, between the two extreme stimuli. And these three groups are arranged, once more, as in Fig. 1. If only a rough approximation to accuracy is wanted, the informal procedure described under minimal changes can be used in evaluating the data. But as has been shown, this procedure is not permissible if it is a question of careful work. Rather after the three groups of values shown in Fig. 1 have been found, these must then be in some so far unexplained way evaluated; in a way, presumably, which will 'approach,' 'resemble' or 'shade off' into one of the telling-off methods about to be described.

In short, mean gradations are no more a method than were minimal changes, but once more a problem. There the problem was to find the threshold of discrimination between sensations; here it is to find the threshold of discrimination between intervals; there the just perceptible difference, here the just equal over-perceptible differences (intervals), between sensations.

Wundt's account of this 'method,' which is doubtless designed to intone the importance of Merkel's 'law,' is peculiar inasmuch as it gives the impression that one must always take the extreme stimuli as fixed and vary the mean stimulus; whereas by far the most work on over-perceptible differences has been done by taking either the upper two or the lower two stimuli as fixed, and the variable stimulus below or above them. When this is done the data still consist in three groups of values for the variable stimulus, according as the variable interval seems greater or less than, or equal to, the standard interval: and these groups still shape themselves as in Fig. 1.

Thus both of the so-called gradation methods are no methods,

but are problems; namely, how to find objective values (measuring the stimuli) which somehow correspond to just perceptible or to equal over-perceptible differences, respectively, between sensations. And either *problem* requires for its solution, firstly, a *procedure* which will yield *three classes of values of the variable stimulus* (corresponding to the judgments 'greater,' 'equal' and 'less') grouped as is shown in a general way by the curves of Fig. 1. Secondly, the problem requires a *procedure* for deriving from these three curves a measure of the discrimination, respectively of difference between sensations or of equality between over-perceptible intervals. The former procedure affords the raw data, the latter evaluates them. In these two processes will be found the real division of psycho-physic methods.

3. TELLING-OFF METHODS: THE METHOD OF MEAN ERROR.¹

This method 'arises from the method of minimal changes in case one limits oneself to taking the just not-perceptible differences of stimuli.' If this were all, then this method would be not an independent method at all, but one of the subdivisions of the so-called method of minimal changes. But it is not all. For while the raw data group once more as in the curves of Fig. 1, we are now for the first time told how to evaluate these data, and to get from the curves a measure of discrimination. The *procedure* becomes here for the first time precise; one finds by definite rules the raw mean error, the variable mean error, the constant mean error; not to mention, as Wundt does not, the probable error, the standard error or deviation, the coefficient of variability, and the relative variability.² The fineness of discrimination is usually taken as the reciprocal of the variable mean error; or, to put it accurately, the discrimination is *defined* as equal to the reciprocal of the variable mean error, called for brevity's sake mean error. By other conventions the other 'errors' are utilized more or less advantageously to give light on the workings of the sense organs under investigation.

¹ Wundt, *op. citat.*, SS. 472-473, 481-482.

² Cf. Yerkes, R. M.: PSYCHOLOGICAL BULLETIN, 1904, Vol. 1, No. 5, p. 137 and Myers, C. S., *Report of the Cambridge Anthropological Expedition*, Vol. II., Pt. II., p. 212.

But this variable mean error measures the sense discrimination not only for the case that one is seeking the just not perceptible difference between stimuli, but also when one wants the just perceptible difference (the quest which is confusingly called the 'method' of minimal changes), or when one wants the just equal over perceptible differences or intervals (called the method of mean gradations). In both these cases also the variable mean error, with the method of finding it, is the precise definition of the terms 'just perceptible difference' and 'equal over-perceptible difference' — terms which were so far loosely used in stating the problems of minimal changes and mean gradations. In short the mathematical manner of evaluating the raw data, which always group as in Fig. 1, is the only exact definition of the measure of discrimination which is being sought.

The method of mean error is then both a problem and a method. Its problem is to find the just not perceptible difference between stimuli. Its method is a mathematical affair of averaging errors and of other details which we are not now concerned with going into; except to note that here at last is a true method. But this method has no intrinsic affiliation with the problem of not perceptible differences: it is equally necessary in evaluating the data on just perceptible differences and just equal over-perceptible differences or intervals. This is a fact which the traditional classification quite obscures, although the actual practice of psychologists will be found to accord with that fact and to belie the tradition.

The three 'methods' of minimal changes, mean gradations, and mean error have amounted so far to *three problems* in three kinds of discrimination, each calling for a special kind of data to be gotten by experiment; and then, *one universally applicable mathematical method* for evaluating these data. The three problems in discrimination may now conveniently be called — just perceptible differences, or *j.p.d.* (from the method of minimal changes), equal intervals that is equal over-perceptible differences, or *e.o.p.d.* (from the method of mean gradations), and not perceptible differences, or *n.p.d.* (from the method of mean error). The one mathematical method for evaluating any of

these data, with its variations, will be called through the rest of this paper the method of mean error, proper. One more *problem* may well be added — that of the threshold of sensation, or *t. of s.* (in comparison with which, as will be recalled, the discriminations just mentioned are often named thresholds of difference).

4. THE METHOD OF RIGHT AND WRONG CASES.¹

This method consists in the repeated presentation to a subject of two stimuli which are so nearly equal that they will frequently be judged to be quite equal, and the greater be sometimes judged even less than the lesser and conversely. Thus the judgments will be sometimes right and sometimes wrong — a fact which every author sapiently points out, as if the same were not true of judgments found under every one of the other

FIG. 2.

'methods.' But the essential fact is that, although these data cannot be grouped like those of Fig. 1, yet they are exactly such a part of Fig. 1 as would lie in an ideally narrow vertical section taken near the zero point of Fig. 1. This is shown in Fig. 2. That is, this method yields as much of the Fig. 1 as a standard stimulus compared with only *one other* stimulus can yield. The judgments gotten are that the compared stimulus

¹ Wundt, *op. cit.*, SS. 473-4, 482-90.

is equal to, greater or less than, the standard; and in order to complete these data into a figure like Fig. 1, it would be necessary only to bring in more compared stimuli, *i. e.*, to use more pairs. This was shown in the second paragraph of this paper. Why sometimes many pair of stimuli are used, and why sometimes only two, does not for the moment concern us.

Now we have seen that the data of Fig. 1 can be evaluated and made to yield a measure of discrimination, by means of a method which we have called the method of mean error. The method of right and wrong cases presents us, in fact, with a *second method* of evaluating data into a measure of discrimination, and relatively fragmentary data at that, *i. e.*, judgments on but two stimuli. It is not the purpose of this paper to discuss these mathematical methods of evaluation in their details; and it is sufficient to recall that a 'Präcisionsmass' is derived from the judgments on two stimuli by means of Gauss's formula for the error curve, or more readily by the use of Fechner's table of integrations derived from that formula. The method assumes that the theory of errors may be applied to the mistakes made in comparing two stimuli, that is, the 'wrong cases'; another fact which is invariably emphasized, although the averages taken in the method of mean error involve exactly the same assumption.

The method of right and wrong cases is seen, therefore, to be no new problem but in fact a new, real method of evaluation, to be used (one is told) when for any reason the stimuli compared have been only two. Right and wrong cases are generally used for finding a 'Präcisionsmass' of the just perceptible difference, but there is no intrinsic reason why the use of two stimuli and Fechner's integral table should not give a measure of equal over-perceptible differences, or even, with the standard stimulus made equal to zero, of the threshold itself.

We have so far analyzed the four so-called psycho-physic methods into four problems (not parallel with the original four 'methods') and two real methods. The problems are those of the *j.p.d.*, *e.o.p.d.*, *n.p.d.*, and that of the threshold of sensation, or *t. of s.* The methods are those of mean error, *m.* of *m.e.*, and right and wrong cases, *m.* of *r.w.c.* Let us

now see what is the relation of these two methods to each other. The *m. of m.e.* is commonly recommended; except where the data are derived from the comparison of only two stimuli, in which case the more cumbersome *m. of r.w.c.* has to be resorted to. But this is not the real distinction between the two methods. It will be recalled by experimenters that the measure of precision yielded by the *m. of r.w.c.* is virtually neither more nor less than the steepness of the curves shown in Fig. 1, or to be more exact, it is a numerical representation of the steepness of the curve of 'greater,' or that of 'less,' judgments (for these two curves are assumed to be ideally similar) when one half of the 'equality' judgments at every point has been added on to this curve (Fig. 3). Now it is always in-

FIG. 3.

sisted that when the judgments are made on but one pair of stimuli, the number taken must be very large indeed if the measure of precision so derived is to be worth anything. But it is seldom if ever suggested, as was done in the second paragraph of this paper, that by the method of least squares the measure of precision may be calculated from more pairs of stimuli and proportionately fewer judgments on each pair. And the results in the two cases will be to all intents and purposes identical. Now this would be applying the *m. of r.w.c.* to the full data of Fig. 1: which shows that the distinction between

the methods of *m.e.* and *r.w.c.* is not one of the meagreness or fullness of the data to be evaluated, that is, is not a question of the number of pairs of stimuli used. But it is a question of the *degree of accuracy* aimed at. The *m.* of *r.w.c.* gives a relatively fine measure of precision, of which the mean error, probable error, and other quantities of the *m.* of *m.e.* are the bare and roughest indications. But so far as these last signify anything, they suggest *approximately* the steepness of that same curve (Fig. 3) of which the 'Präcisionsmass' of the *m.* of *r.w.c.* is a relatively *accurate* measure. The difference between the two mathematical methods is thus purely one of degree of accuracy, and it is a mere accident of technique that the *m.* of *m.e.* cannot be used with judgments on only one pair of stimuli, nor the *m.* of *r.w.c.* for the determination of not perceptible differences. In fact the latter is possible if an experimenter should choose to neglect a part of the wrong cases, taking only those which form the curve of 'equal' judgments (Fig. 1), and deriving the steepness of this curve by means of Fechner's table. For this would be a permissible variation of the *m.* of *r.w.c.*, comparable to the several variations in the *m.* of *m.e.* which have been used and recommended.

It may well be questioned whether the *m.* of *r.w.c.* is not a much more accurate procedure than the conditions of experimentation ever justify; or whether the application of least squares would not be a case of penny-wise after pound-foolish, in any sort of physiological work whatsoever. But we are not here concerned with the mathematical minutiae of method, — only with the general classification. The new tent poles are to carry all the old canvas. Indeed if mathematical details were here in question, it would be our first duty to examine and if possible to justify the fundamental assumption of both methods, *i. e.*, that the theory of errors may be applied to curves which never are and by all psycho-physic laws never can be, truly symmetrical.

To survey our results once more, the four psycho-physic 'methods' resolve themselves into the four problems, of finding the threshold of sensation, and the thresholds of not perceptible, just perceptible, and equal over-perceptible, difference; and

then the real two methods of mean error and right and wrong cases. The four original 'methods' are curiously illogical classes, which have come about in the course of the development of psycho-physics. They are historical relics. The first two 'methods' are only problems; the third is a problem and a method; while the fourth is a method but not a problem.

We have seen that in all psycho-physical experimentation there are two stages of the work — the getting of data, and the evaluating of them. We have already considered, so far as it lies in our purpose, the second, purely mathematical stage. It is instructive, and in part will account for the retention of the false methodology, to note how the experimenter has to approach the first stage of his work, the getting of data. Suppose that he wants to study by means of the ordinary olfactometer and one of the four 'methods' the discrimination for odors. He cannot use the 'method of mean gradations' because few if any subjects are able to identify and hold in mind an over-perceptible difference between odors. So the experimenter thinks that he must choose another 'method'; the fact is that he must choose another problem, that is, he must not hope to measure the olfactory discrimination for over-perceptible differences. Similarly he will believe himself deterred from the 'method of mean error,' because he has been taught that this 'method' involves the adjustment of the stimuli by the subject himself: and of course an odor-tube cannot be adjusted back and forth like a monochord. He may think of doing himself, as experimenter, the adjusting, continuing each time, although this is contrary to the school directions, until the subject declares a just not perceptible difference. This is possible to do, but hardly advisable; since it happens by an accident of physics that the olfactometer is more suitably and naturally adjusted from less olfactory stimulus to more, rather in the opposite direction. It is an accident, then, that the experimenter cannot well study just not perceptible differences, but must choose the remaining problem, the 'method' of just perceptible differences.

It may be said that this necessity of casting about for a problem (*n.p.d.*, *j.p.d.*, *e.o.p.d.*) suited to the accidental pecu-

liarities of the sense-organ to be studied and the apparatus at one's disposal, has made the spurious distinction between 'methods' seem real and practical to experimenters who seem not to have noticed that the distinction is in problem and not in method.

Thus the first stage of psycho-physical experimentation, the getting of data, is practically not determined by the experimenter, but by relatively accidental circumstances, — peculiarities of the sense-organ chosen and of the apparatus available. And in determining what data can be gotten, these accidents determine at the same time what problem can be studied, that is, whether *n.p.d.*, *j.p.d.*, or *e.o.p.d.* The word 'accident' may be objected to. Yet it is fair to call accidental the circumstance, for instance, that the *n.p.d.* cannot be studied in the active muscle-sense. Such seemingly chance and irrelevant factors are almost numberless. An important one of them, although it influences the general problem less directly, is the possibility of individual or massed stimulation of end-organs. By an anatomical accident the olfactory end-organs cannot be stimulated individually; so that what is an interesting problem of the dermal senses, the relative thresholds of different individual end-organs, cannot be studied in the sense of smell. It is apparently an accident, though an interesting one, that some sense-organs, as the olfactory, gustatory, or dermal when stimulated singly, do not afford us clear sensations of intervals, *i. e.*, of over-perceptible differences. Hence their power of discrimination must be studied in other respects; the problem of *e.o.p.d.* is debarred. Once again, if the experimenter happens to have two tuning-forks and nothing else, he will necessarily adapt his problem to the *m.* of *r.w.c.* with which he will have to evaluate his data. Whereas if he had a sonometer, he would more naturally let the subject adjust for himself and give judgments of *n.p.d.* The case in which the investigator has but two different stimuli to apply (and these must be nearly alike) is the one case in which truly the *method as well as* the problem is determined by irrelevant circumstances. Otherwise the *method* is chosen (*m.* of *m.e.* or *m.* of *r.w.c.*) according to the *degree of accuracy* which is desired.

Aside from the accidents which determine the problem, there is another kind which influences only the technique or at best bears but remotely on the problem. Such a factor, for instance, is the matter of simultaneity or succession of stimuli. Since the olfactory end-organs have to be stimulated all at once, any discrimination which is studied has to be a successive one (apart from the very doubtful case which some would claim, of simultaneous bilateral stimulation). This circumstance would still leave as possible any of the three problems; but it helps to restrict more precisely the technique to be adopted. There are countless other accidents of a similar sort.

Therefore in first approaching a psycho-physical problem, and in trying to get it realized in some arrangement of apparatus, the experimenter finds that very little is left to his free choice. Sometimes indeed he may choose no more than barely the sense which he studies. Two kinds of accidental circumstances, as we have seen, restrict his course, although in practice it is scarcely necessary to distinguish between them. In order to accept the inevitable, and set up his apparatus with the least waste of time and thought, the experimenter must run through the possible limitations and find out definitely what ones actually confront him. The possibilities are fairly well included under the following categories, although the list aims merely at being serviceable but not exhaustive.

1. Comparison — (*a*) simultaneous, (*b*) successive.
2. Comparison — (*a*) direct (immediate), (*b*) mediate.
3. Comparison between — (*a*) two stimuli, (*b*) more than two stimuli.
4. Variation — (*a*) random, (*b*) progressive.
5. Variation — (*a*) continuous, (*b*) discontinuous (step-wise).
6. Apparatus operated by — (*a*) the experimenter, (*b*) the observer.
7. Actual relation of the stimuli — (*a*) known to the observer, (*b*) not known to the observer.
8. (If the sense to be studied and the above enumerated conditions are such as still to leave the question open): Discrimination of — (*a*) threshold of stimulation, (*b*) not perceptible difference of stimulation, (*c*) just perceptible difference of stimulation, (*d*) equal over-perceptible differences of stimulation.

These headings are all familiar to the psychologist and need no elucidation. They are neither completely independent nor yet mutually exclusive. Class 1, for instance, is independent of 2; but 3, *a*, excludes 4 and 5. After learning what of these alternatives are open, the experimenter will see his way of proceeding rather precisely marked out. Herewith the first methodological stage is ended. Before coming to the second stage, of mathematical evaluation, he has only to get his data.

In regard to this second stage we have already seen that there are only two ways of evaluating data (although each method allows some minor variation), the methods of mean error and of right and wrong cases. It was not the purpose of this paper to discuss the methods in detail, but only to analyze the so-called four 'methods' and to classify the results. This we have now done. It need only be noted once more, that the choice between the two actual methods (except in the case of only two stimuli being used) depends on nothing but the degree of accuracy which is desired, that is, on the amount of labor which the experimenter thinks proper to devote to the inquiry. We may now pass to two of the other emendations of the traditional methodology, which have been offered.¹

¹The writer greatly regrets that before this article was actually set up, he had not seen the admirable work of G. E. Müller, in the *Ergebnisse der Physiologie*, 2ter Jahrgang, 1903, II Abth., SS. 266-516. Müller finds four cases (Fälle) in which psycho-physic methods may be used. These are our *t. of s.*, *j.p.d.*, *n.p.d.* and *e.o.p.d.* Of 'methods' he finds three; the first, in which the observer adjusts the variable stimulus in random sequence; the second, in which the experimenter does this, but in orderly increasing or decreasing progression; and the third, in which the stimuli are not adjusted (old method of right and wrong cases). There then remains the task (Aufgabe) of finding in general two values, a mean and its variability. This may be done immediately (our *m. of m.e.*), or by the mediation of formulæ (our *m. of r.w.c.*).

Of course this is in essentials far nearer to that which has been urged above than is any other classification hitherto offered. To the present writer it still seems, however, that the 'cases' are problems of which accidental circumstances largely determine the choice; that Müller's 'methods' are merely three among a large number of equally important such accidents; and lastly, that the actual methods are Müller's two Aufgaben, the treatment of the results either with or without formulæ. But only the second point is a difference in principle, while the first and last are merely nominal.

THE CLASSIFICATION OF KÜLPE.¹

Külpe's analysis of the methods is not fundamentally different from that of Wundt, but there are superficial differences which are worthy of a brief consideration. For Külpe there are two groups of methods, that of minimal changes and that of errors; these two correspond to Wundt's gradation and telling-off methods. The method of minimal changes has four 'applications'; while the error methods are two—the method of right and wrong cases and the method of mean error.

The method of minimal changes may be 'applied' to the determination of threshold (*Reizbestimmung*), to the comparison of stimuli (*Reizvergleichung*, the *n.p.d.* mentioned above), to the determination of difference (*Unterschiedsbestimmung* or *j.p.d.* above), and to the comparison of differences (*Unterschiedsvergleichung* or *c.o.p.d.* above). This virtually admits, although Külpe seems unconscious of the fact, that these four groups are not methods but problems, as has been argued in this paper. Furthermore his classification is symmetrical and consistent: *i. e.*—

1. Threshold—(*a*) of sensation, (*b*) of interval between sensations.

2. Equality—(*a*) of sensations, (*b*) of intervals between sensations.

Külpe well says that there is one method, which he calls the 'method of minimal changes, that applies to these four classes. This 'method' is essentially like the procedure described by Wundt under the same name, and is subject to the objection which was noted in the early part of this paper. This is, as will be remembered, that whether the compared stimulus varies continuously or step-wise, there is no point at which the change in judgment from being always 'greater' or always 'less' to being for the first time 'equal,' or vice versa, is truly significant. For let us suppose the compared stimulus to be decreasing toward the standard, it may be considerably greater than this and be judged 'equal' while when further decreased it will be again judged 'greater.' Or if it is increasing to the standard, it will often be judged 'equal' when considerably less than the

¹ Külpe, O., *Grundriss der Psychologie*, Leipzig, 1893, SS. 55-81.

standard, but again 'less' when it has been increased nearly to equality with the standard. To ignore these inconsistencies and to interrupt the comparisons with the first judgment of equality which is given (or inequality, as the case may be) is, if the compared stimulus varies step-wise, to take a measurement of discrimination which depends materially on the size of the steps which have been arbitrarily chosen. If the compared stimulus varies continuously, the case is a trifle better but not much, since then the measure of discrimination is considerably vitiated by fatigue and expectation (see above). Furthermore it is often not possible to make the compared stimulus vary continuously.

For these reasons Wundt virtually admits, as we have seen, that here is no method: and of his 'method of minimal changes' he leaves, although merely by implication, only the problem of the *j.p.d.* Külpe, however, accepts the informal procedure disparaged by Wundt, and insists that it is a method. It has been asserted elsewhere in this paper that every psycho-physical method must involve a procedure which yields at least a part of the data shown in Fig. 1, and must then mathematically evaluate from these some sort of a measure of discrimination. Now what part of Fig. 1 does this procedure yield which Külpe recommends? And how are these data evaluated? The 'method' is to present to the subject a pair of stimuli, a standard and a compared, for his judgment of 'less,' 'equal' or 'greater.' The compared stimulus varies in successive presentations not at random but so as to approach or depart from the size of the standard stimulus; and it may do both, either above or below this standard. Thus there are four modes in which the compared stimulus can vary. Whether in a given experiment some or all of these modes are used depends on the problem to which this 'method of minimal changes' is being applied.

Let us suppose, for example, that it is being applied to the problem of *j.p.d.* All four modes are used. The compared stimulus (*c.s.*) starts so much smaller than the standard stimulus (*s.s.*) that it is always judged smaller; and is gradually increased until the first judgment of equality is gotten. Here the progressive comparisons are interrupted, although if *c.s.* were

further brought up toward *s.s.* it would very likely be judged less once or twice more. Here is of course the weakness of this method. Now *c.s.* is taken equal to *s.s.*, and made gradually to decrease until the first judgment of less is given. So far the experimenter has a series of judgments 'less' 'less' et cet. — 'equal,' for *c.s.* increasing; and a series 'equal' 'equal' et cet. — 'less,' for *c.s.* decreasing. In both cases *c.s.* is smaller than *s.s.* But it can increase and decrease while larger than *s.s.* Therefore two more similar series are gotten, for values of *c.s.* above *s.s.* Curiously enough, Külpe writes as if it were enough to take the average of only these four values of *c.s.* at which the first change in the judgment occurred.¹ But he can

FIG. 4.

not mean so, for it appears later,² that a mean variation may be in question; so that of course, as one would expect, the four series are to be gotten many times, and an average of all taken for the actual measure of discrimination. The reader will see at once that if plotted these raw data would look like Fig. 4.

The relation to Fig. 1 (the dotted lines) is clear enough. Where the differences between *c.s.* and *s.s.* are so very small or so very large as always to be correctly told, the curves run smoothly. They become jagged where the chance errors interrupt the various series of 'less,' 'less,' et cet. —, 'equal,'

¹ Külpe, *op. cit.* SS. 60-62.

² *Ibid.*, SS. 66.

'equal' et cet.—, or 'greater,' 'greater' et cet.—, with a change in the judgment uttered. Now, as we have already seen, it is just because the first change in judgment depends more on the size of the step variations of *c.s.*, on fatigue or expectation, than on the fineness of discrimination in question that this method, as even Wundt grants, is not a proper method. It selects for further mathematical evaluation a certain portion of the data of Fig. 1; but the principle of selection is largely, though not utterly, independent of the power of discrimination which is being studied. And whereas in the *m.* of *r.w.c.* when only two stimuli are used, the selection although arbitrary yet leaves enough of the essential parts of the curves; the principle of selection here used effectually excludes just these indispensable parts, namely the parts which show the characteristic steepness of the curves.

Therefore in the earlier part of this paper the name of method was justly denied to this procedure, since it is absolutely inadmissible. And it was insisted that the only allowable and yet mathematically convenient method was to take data giving the complete curves of Fig. 1, and then to express approximately the steepness of those curves (transformed perhaps as in Fig. 3) by means of a mean error, probable error, or by some such readily obtainable quantity. Külpe's method of minimal changes is simply under no circumstances allowable. The merit of this part of his classification lies wholly in his four-fold 'application' of the 'method of minimal changes.'

The other part of his classification needs but brief consideration. As to his 'method of right and wrong cases,' he well says:¹ That it 'is capable of as manifold application as the method of minimal changes'; but then he goes on to describe the method so wholly in the traditional way that one has the impression that the use of only two stimuli is as essential to this method as is Gauss' equation itself.

Külpe's treatment of the 'method of mean error' adds nothing to that of Wundt, except the statement that while the method has been used only for the *n.p.d.*, it could be used for the *e.o.p.d.* as well. But he declares (S. 78) that it can be used

¹ *Ibid.*, S. 70.

only when experimenter and subject are one person, and when the stimulus can vary continuously. Both of these statements we have seen to be untrue.

THE CLASSIFICATION OF EBBINGHAUS.¹

The other classification which will concern us in this paper is that of Ebbinghaus, given after his short and admirable account of the traditional four methods. Ebbinghaus says (S. 75) that, 'The goal of psycho-physic method is the determination of those stimulation values which are the outward causes of equal-seeming psychic values (that is, of equal intervals between sensations),' of course, then, the threshold of stimulation is not a problem of psycho-physics; and in fact the author makes no mention of it.

Now as to the size of the intervals between sensations, it can be either just perceptible or over-perceptible, and if the latter it can be as large as one chooses. Herewith our problems of *j.p.d.* and *e.o.p.d.* are recognized, although still called methods; while the problem of *n.p.d.* is ignored. Secondly, as to the way of judging, this may be so chosen that the subject (S. 75) 'has in mind the idea of an equal interval and varies the outward stimulus until the sensation which it gives corresponds to this idea: or, one can present repeatedly for judgment a given pair of stimuli and let the subject give his judgment in terms of certain replies previously agreed on,' such as 'less,' 'equal,' and 'greater.' The first procedure is called the 'finding a stimulus' to correspond to a given judgment; the second, the 'finding of judgment' to correspond to a given pair of stimuli. Lastly, in every case the final result must be the average of many observations; and since each observation will vary from this average, the mean of the variations must always be given (mean deviation), in order to show the reliability of the total average.

This scheme is very original and suggestive; but inadequate. It recognizes but one true method, the one which we have called above the *m. of m.c.*; and but two problems (which Ebbinghaus still calls methods) to which this one method can

¹ Ebbinghaus, H., *Grundzüge der Psychologie*, Leipzig, 1902, SS. 74-80.

be applied; those are our problems of *j.p.d.* and *c.o.p.d.* The problem of threshold of stimulation is dismissed as not lying within the field of psycho-physics. The expediency of so narrow a definition of psycho-physics is doubtful; for it should seem better to account the problem of psycho-physics the measured correlation in general of stimulus and sensation; and thus to include the problem of threshold, which after all must be studied if at all by the *m.* of *m.e.* or that of *r.w.c.* Also the problem of *n.p.d.* finds no recognition because, as the author says (S. 78): 'The third method, finally, that of mean error [that is the third traditional method, by which he refers to the problem of *n.p.d.*], affords nothing directly which can be utilized as a measure of sensation, since it operates not with a sensation interval, but with the disappearance of an interval'; *i. e.*, it operates with a not perceptible interval, which is seemingly not to the point. And again he says (S. 69): 'The determination of this mean error [made in judgments of *n.p.d.*] through the range of a given sense has of course its use, but clearly these values are something quite different from equal sensation intervals in the above mentioned sense [that is, of just perceptible, and equal over-perceptible, intervals], and the process of getting them is no true measuring of sensation but a process having some relation to such a measuring.' Ebbinghaus gives elsewhere his grounds for this view, in that the *n.p.d.* never had any value except on the assumption that it bears a definite and fixed relation to the *j.p.d.* from the same standard stimulus. This assumption has not been shown to be valid, and Ebbinghaus seems to account it very speculative if indeed not certainly false. He may be quite right. The point is one more of mathematical technique than of classification. If he once granted the validity of the *n.p.d.*, Ebbinghaus would doubtless class it with the *j.p.d.* and *c.o.p.d.*

The analysis given in this paper found the third so-called 'method' of tradition to consist in a problem and a method. Ebbinghaus discards the problem but retains the method, which is our *m.* of mean error. This is in fact his one method, and it must be used in every one of the four procedures into which he resolves the traditional 'methods.' These are, once more:

1. Just perceptible interval with finding of stimulus.
2. Just perceptible interval with finding of judgment.
3. Over-perceptible interval with finding of stimulus.
4. Over-perceptible interval with finding of judgment.

Thus in every psycho-physical procedure an average of the observations must be taken, with their mean variation or error. These two quantities are the measure of discrimination which is desired.

But what of the 'Präcisionsmass' as yielded by Gauss's equation and Fechner's table of integrals? Ebbinghaus minimizes in his system this method of evaluation because, it should seem, he accounts it a very laborious process to be used only when all the observations are based on but two stimuli: the more cumbersome mathematical evaluation being used solely in order to make up for the poverty of the material of observation. The author does not admit, apparently, that here is an evaluation method of relatively great accuracy, which can be applied as well to observations on many pairs of stimuli as to those on only one pair.

We have seen that from the first and second traditional methods Ebbinghaus analyzes out his two categories of just perceptible and over-perceptible, difference. Both of these 'methods' proceed by '*finding the stimulus*,' whereas the fourth 'method' (of right and wrong cases) proceeds by '*finding the judgment*'; hence these two new categories. Every psycho-physical procedure uses one of the last two, together with one of the first two categories. Thus arise the four methods of Ebbinghaus, given on the preceding page. But now all four of these methods have to use the little germ of truth which lay in the traditional third 'method' (of mean error): this was the use of the average and its mean error. Wherefore the final form of Ebbinghaus's four methods is as follows:

1. Just perceptible interval with finding of stimulus; average of all observations with *m.e.*
2. Just perceptible interval with finding of judgment; average of all observations with *m.e.*
3. Over-perceptible interval with finding of stimulus; average of all observations with *m.e.*

4. Over-perceptible interval with finding of judgment ; average of all observations with *m.e.*

Thus this system, which is somewhat simpler than the traditional four, is gotten by the omission of two problems, those of the threshold of stimulation and of the *n.p.d.* ; and by the omission of one method, that of right and wrong cases. It is curious that aside from the relatively special and insignificant expedient for evaluating judgments on only two stimuli, Ebbinghaus finds nothing in the traditional method of right and wrong cases save the suggestion to fit the judgment to the stimulus instead of the stimulus to the judgment, as in minimal changes.

Although this classification leaves out so much that it becomes inadequate to the subject, it probably has the distinguished merit of being the first radical and strictly logical recasting of the methodology. It is well worthy of study, and will be found to be possibly, the clearest and best form for teaching the methods of psycho-physics to elementary students.

SUMMARY.

The four traditional methods of psycho-physics are found to be an illogical scheme of a subject which if analyzed resolves itself into the following system.

I. Four problems as to the relation between stimulus and sensation, which admit of quantitative treatment.

- (a) The threshold of stimulation (*t. of s.*).
- (b) The not perceptible difference of stimulation (*n.p.d.*).
- (c) The just perceptible difference of stimulation (*j.p.d.*).
- (d) The equal over-perceptible difference of stimulation (*e.o.p.d.*).

II. A great diversity of procedure by which data on these problems are gotten (see above), and of which the one thing essential is that judgments shall be obtained which group themselves like a part or like the whole of the judgments in Fig. 1, and so that the steepness of at least one of the curves of Fig. 1 is implicitly contained in the judgments. This variety of procedure is not a variety of choice open to the experimenter. The procedure used in any case depends on relatively acci-

dental characteristics of the sense-organ chosen for study; and for this reason they are not susceptible of rigorously logical classification. Furthermore it generally happens that these different accidents, so far from being a source of freedom, actually restrict the experimenter in his choice of problem, so that he is obliged to study a particular one of the four problems.

III. Two real methods by which the judgments can be evaluated into a measure of the discrimination of the sense studied: that is, by which the steepness of one or all of the curves in Fig. 1 can be approximately or accurately expressed. These may be called, out of respect to tradition.

(a) The method of mean error (*m. of m.e.*).

(b) The method of right and wrong cases (*m. of r.w.c.*).

The difference between these is one of accuracy. The former method measures the precision roughly by taking an average of the observed values and their mean error, or probable error, et cet. The latter method is more refined and uses the equation of Gauss and the integral tables of Fechner to obtain a 'Präcisionsmass.' The significance of either of these measures lies in its being an index of the steepness of one or all of the curves of Fig. 1.

Either of the two methods may be used with any of the four problems. There are thus eight alternatives. Having chosen the sense-organ which he will study, the experimenter finds his mode of procedure limited in many respects by accidental peculiarities of this sense and of the apparatus at his disposal. He may then choose what of the four problems he will study, or he may find even this determined by the accidental peculiarities. He is, however, free to choose his method of evaluation: and the basis of his choice is nothing but the degree of accuracy which he desires in his results or, what generally comes out to the same thing, the amount of labor which he is willing to spend on them. The one exception is the case in which some circumstance limits him to the use of only two stimuli, for then it limits him also to the *m. of r. w. c.*¹

¹ The MSS. of this article was received April 17, 1904.—ED.