PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A
MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

On the Mathematical Theory of Errors of Judgment, with Special Reference to the Personal Equation

Karl Pearson

Phil. Trans. R. Soc. Lond. A 1902 **198**, 235-299
doi: 10.1098/rsta.1902.0005

V. *On the Mathematical Theory of Errors of Judgment, with Special Reference to the Personal Equation.*

By KARL PEARSON, *F.R.S., University College, London.*

Received April 23,—Read June 20, 1901.

CONTENTS.

Introductory.

1. THE following investigation has been in progress for some years and led to a paper, communicated to the Society on December 29, 1896.* I therein pointed out that personal judgments were frequently correlated. This correlation may be of the kind which in that paper I termed "spurious," or it may be genuine. By "spurious" correlation I understand the quantitative measure of a resemblance in judgments, which resemblance is due solely to the particular manipulation of the observations. Very customary treatment of observations will lead to the existence of a spurious correlation, which may be and generally is entirely overlooked by the observers. For example: if the quantity to be determined by judgment were the time taken by a bright point, say a star, in travelling from a position C intermediate between spider lines A and B to the line B, and the result were to be expressed by the ratio of this time to the known time from A to B, then there would be correlation in the results obtained by two observers for a number of stars, even if their absolute judgments on the

* 'Roy. Soc. Proc.,' vol. 60, p. 489.

time from C to B were quite independent. Again, if the judgments of two observers be in both cases referred to a standard observer, then such relative judgments will be found to be correlated; and this is true, although if we could find the *absolute* errors of the two observers, we might discover that these errors were quite uncorrelated. We shall see illustrations below of the manner in which this spurious correlation almost imperceptibly creeps into any ordinary method of manipulating observations, and how very little attention has hitherto been paid to it.

But apart from this spurious correlation the experiments described in this memoir seem to show that there exists almost invariably a genuine correlation between the judgments of independent observers. This may be due to two sources: (i.) Likeness of the environment in the case of each individual observation, which leads to likeness of judgment in the individual observers. One experiment may appear to be made under precisely the same conditions as a second, but really it has a certain atmosphere of its own which influences the observers in a like manner. (ii.) Likeness in the physical or intellectual characters of the observers leading to a likeness in their judgments of what took place.

It is usual to suppose that the error made by an individual observer depends upon a great variety of small causes largely peculiar to that individual; or, if peculiar to the individual experiment, that they will affect different observers in different ways. Our experiments show such considerable correlation between the judgments of individual observers, that I have been compelled to discard this view; I consider that very slight variations of the environment (for example, similar observations on stars of different N.P.D.) will be quite sufficient to produce correlated judgments; on the other hand, some slight similarity of eye-sight, of ear, of temperament, may be sufficient to associate the judgments of two observers. Whatever variety of small causes influence the judgment, it is clear that in actual practice they do not suffice to dominate some particular source of mental or physical likeness which leads to this correlation in judgments.

Our first series of experiments show that the actual instantaneous environment is not necessarily the source of likeness in judgment. The same lines were not dealt with by the three observers at one and the same instant. Thus it is on some quite definite, but probably quite undiscoverable, likeness of temperament that we must largely rely to account for this correlation of judgment.

To the naturalist, who has to observe, whether he be physicist, astronomer, or biologist, this genuine correlation of judgments is of equal significance with the "spurious" correlation, and, like the latter, almost invariably disregarded. A and B are two independent observers, making an experiment of the same character, or observing the same phenomena. As a rule their judgments, however, will not be independent. The importance of this conclusion in modifying the weight which must be given to a series of observations of the same phenomena made by two "independent" observers will be manifest. Once we admit that the judgments of independent observers are correlated, then the determination of the amount of correlation becomes of vital importance.

It has only been after further experiment, and after much seeking for possible sources of spurious correlation, that I have at last convinced myself of the reality of this genuine correlation in the judgments of independent observers. I cannot expect my readers to do so at once, but I believe that a careful examination of our experimental results will at least convince them that it is a factor of great importance in some, if not, as I believe, in all types of observation. As to the spurious correlation, it plays such a large part in relative personal judgments, and is so obvious from the theoretical standpoint, that one can only wonder it has not hitherto been regarded.

The course which I propose to follow in this memoir may be thus summed up :—

(*a.*) I shall introduce a more complete terminology than appears at present to exist for the theory of errors of judgment.

(*b.*) I shall develop to some extent the current theory of errors, and its application to personal equation.

(*c.*) I shall next consider what modifications must be made in this theory to allow for the correlation of the judgments of independent observers.

(*d.*) I shall then discuss certain experimental investigations on personal equation, which demonstrate that (*c.*) and not (*b.*) is the category under which we must class errors of judgment.

(*e.*) Lastly, I shall sum up the bearing of this discussion on our treatment of errors of observation, whether physical or astronomical.

(2.) *Terminology.*

If ξ be the actual value of some physical quantity, whether it can be really determined or not, and x_1, x_2 be the values of it according to the judgments of two independent observers, whether formed by measurement, estimate, chronographic record, or any other way, we shall speak of $x_1-\xi$, $x_2-\xi$ as the absolute errors of judgment of the two observers. x_2-x_1, which in many cases is all we can determine, will be termed the relative error of judgment of the two observers.

If a sufficiently large series of judgments be taken, then the mean values of $x_1-\xi$ and $x_2-\xi$ will be termed the absolute personal equations of the observers, and the mean value of x_2-x_1 the relative personal equation of the two observers. We shall use the notation p_{01}, p_{02} for the absolute, p_{21} for the relative personal equations of the two observers.

Clearly $$p_{21} = -p_{12}.$$

If we form the standard deviations of the absolute judgments σ_{01} and σ_{02}, and of the relative judgments $\sigma_{21}=\sigma_{12}$, these will be measures respectively of the variability in judgment of either observer absolutely, and of the variability of their relative judgment.

We have, if n be the number of judgments in the series :—

$$\left.\begin{aligned}\sigma_{01}{}^2 &= \frac{1}{n}\,\mathrm{S}\{p_{01} - (x_1 - \xi)\}^2,\\ \sigma_{02}{}^2 &= \frac{1}{n}\,\mathrm{S}\{p_{02} - (x_2 - \xi)\}^2,\\ \sigma_{21}{}^2 &= \frac{1}{n}\,\mathrm{S}\{p_{12} - (x_2 - x_1)\}^2\end{aligned}\right\} \quad . \; . \; . \; . \; . \; . \; . \quad \text{(i.)},$$

where S denotes a summation for every judgment of the series.

Obviously the goodness of an observer is measured by two characters :

(i.) The smallness of his personal equation, p_{01}.

(ii.) The smallness of the variability of his judgment, σ_{01}.

The first determines the average error of his judgment, the second the constancy or stability of his judgment.

The latter is often quite as important a feature of the mental worth of an observer as the former.

This steadiness or reliability of judgment, which I shall term stability of judgment, will be defined as follows :—The relative stability of two observers for a given class of observations is measured by the inverse ratio of their standard deviations. Or, if we are speaking of the same class of observation the absolute stability of judgment is $\frac{1}{\sigma_{01}}$. In the case of relative judgments, $\frac{1}{\sigma_{21}}$ will measure the steadiness in relative appreciation of two observers ; it serves as a measure of their degree of approximation to like estimates, and may be called their relative stability. It by no means follows, however, that two observers with a large degree of relative stability have necessarily large individual absolute stabilities in judgment, nor that their absolute personal equations are small. This remark is of considerable importance, for we are apt to think that if two out of three observers have a small relative personal equation and a large relative stability, then their conclusions are worth more than those of a third observer with whom they have large relative personal equations and smaller relative stabilities.

No conclusion of this kind can be admitted, if we find that the absolute judgments of independent observers are correlated ; for, as will be shown later, the higher this correlation, *i.e.*, the less independence in judgment, the greater becomes the relative stability of the two observers. The more marked this association in judgment, the less are we able to set the judgment of two observers against a third.

The correlation in absolute judgments between two observers* is given by

$$r_{12} = \frac{\mathrm{S}\{(p_{01} - (x_1 - \xi))\,(p_{02} - (x_2 - \xi))\}}{n\sigma_{01}\sigma_{02}} \quad . \; . \; . \; . \; . \; . \quad \text{(ii.)}.$$

* 'Roy. Soc. Proc.,' vol. 60, p. 480 *et seq.*

The correlation in the relative judgments of two observers, 1 and 2, both referred to a standard observer 3, is given by

$$\rho_{3,\,12} = \frac{S\{(p_{31} - (x_3 - x_1))\,(p_{32} - (x_3 - x_2))\}}{n\sigma_{31}\sigma_{32}} \quad . \quad . \quad . \quad . \quad . \quad . \quad (\text{iii.}).$$

So far as I can make out it is usually assumed that r_{12} is zero, and the existence, if r_{12}, &c., be zero, of very sensible values in the case of $\rho_{3,\,12}$ has always been disregarded.

The probable errors* of personal equations, variability in judgments, and correlations in judgments, as determined by the formulæ (i.) (ii.) (iii.) above, are :—

$$\left.\begin{array}{lll}
\text{Per cent. of } p_{01} & = & \cdot 67449\ \sigma_{01}/\sqrt{n} \\
\quad ,, \quad ,, \ p_{02} & = & \cdot 67449\ \sigma_{02}/\sqrt{n} \\
\quad ,, \quad ,, \ p_{21} & = & \cdot 67449\ \sigma_{12}/\sqrt{n} \\
\quad ,, \quad ,, \ \sigma_{01} & = & \cdot 67449\ \sigma_{01}/\sqrt{2n} \\
\quad ,, \quad ,, \ \sigma_{02} & = & \cdot 67449\ \sigma_{02}/\sqrt{2n} \\
\quad ,, \quad ,, \ \sigma_{21} & = & \cdot 67449\ \sigma_{21}/\sqrt{2n} \\
\quad ,, \quad ,, \ r_{12} & = & \cdot 67449\ (1 - r_{12}^2)/\sqrt{n} \\
\quad ,, \quad ,, \ \rho_{3,\,12} & = & \cdot 67449\ (1 - \rho_{3,\,12}^2)/\sqrt{n}
\end{array}\right\} \quad . \quad . \quad . \quad . \quad (\text{iv.}).$$

If any investigation of personal equation is to have validity these probable errors must be small relatively to the quantity measured. Accordingly, no determination of personal equation is of the slightest value which does not give σ as well as p, for without this we do not know the weight to be attributed to the determination of p. My own experience would seem to show that ten to thirty observations, on which number some estimates of personal equation have been formed, are very insufficient. Further, astronomers rarely publish the data on which the personal equation has been determined so as to enable one to judge of its degree of stability, or of the degree of independence in the judgments of different observers.

We shall have to investigate whether there are methods of finding σ_{01} and σ_{02} when only the relative personal equations and relative variabilities are given, and we shall have to see how the correlation of absolute and relative judgments may be determined.

Personally it appears to me that without a knowledge of all these quantities we cannot profitably combine the observations of different observers or determine their individual independence and stability of judgment.

* 'Phil. Trans.,' A, vol. 191, pp. 239–245.

(3.) *Current Theory of Errors of Observation.*

The assumption usually made is that the error of an observation is due to the result of the combined action of a great number of independent sources of error; each source follows a permanent law and attributes equal probability of occurrence to numerically equal errors. From this statement, or some modified form of it,* is deduced the well-known normal curve of error frequency :—

$$y = y_0 e^{-x^2/2\sigma^2} \quad \text{(v.)}.$$

An important point to be considered is, therefore, whether actual errors of observation in any case are such that they may be supposed to be a random sampling of errors obeying this law. I have in a recent paper† obtained a criterion for the probability of any system being the result of a random sampling from a series following any law of frequency, and I have shown that it is most highly improbable that the series cited by AIRY and MERRIMAN as evidence of the suitability of the normal curve can really have been random samples from material actually obeying such a distribution.

Assuming the applicability of the normal curve, or, indeed, the independence of judgments of independent observers,‡ we have at once

$$\left.\begin{aligned} &\sigma_{21}^2 = \sigma_{01}^2 + \sigma_{02}^2, \\ \text{Similarly :—} \quad &\sigma_{32}^2 = \sigma_{02}^2 + \sigma_{03}^2, \\ \text{and,} \quad &\sigma_{13}^2 = \sigma_{03}^2 + \sigma_{01}^2 \end{aligned}\right\} \quad \text{(vi.)}.$$

Hence we deduce :—

$$\left.\begin{aligned} \sigma_{01}^2 &= \frac{\sigma_{21}^2 + \sigma_{13}^2 - \sigma_{23}^2}{2}, \\ \sigma_{02}^2 &= \frac{\sigma_{32}^2 + \sigma_{21}^2 - \sigma_{13}^2}{2}, \\ \sigma_{03}^2 &= \frac{\sigma_{13}^2 + \sigma_{32}^2 - \sigma_{21}^2}{2} \end{aligned}\right\} \quad \text{(vii.)}.$$

It was this simple result which led to the whole of the present investigation. I had not seen it noticed before, and it seemed of wide-reaching importance. I mean in the following manner: The astronomer, and often the physicist, can, as a rule, only determine relative and not absolute judgments. He cannot deduce the absolute

* These are really additional assumptions. See pp. 274–275 later.

† 'Phil. Mag.,' July, 1900, p. 157 *et seq.*

‡ If z_1 and z_2 be judgments of two observers and z_{12} their relative judgment, δz_1, δz_2, δz_{12}, errors measured from the means of the respective systems, then $\delta z_{12} = \delta z_1 - \delta z_2$, whence the result follows at once, if the correlation $= \dfrac{S(\delta z_1 \times \delta z_2)}{n\sigma_{01}\sigma_{02}}$ be zero.

personal equations from his knowledge of the relative personal equations. How then is he to measure the relative goodness of observers? In turning this problem over in my mind it occurred to me that if there were no means of measuring the average absolute error of an observer short of an experiment *ad hoc*,* still, if we could deal with *three* observers, their relative variabilities would give us the means of determining their absolute variabilities, and the astronomer or physicist would thus really be in a position to judge something about the steadiness in absolute judgment of a series of observers. He could find their σ_{01}, σ_{02}, and σ_{03}, and so determine their stabilities.

Now, if one accepts the independence of the judgments of independent observers, (vii.) follow at once, and we have an important problem simply solved. I therefore organised a series of experiments to illustrate (vii.), but instead of discovering a new method of testing observers' stability of judgment, I found that (vi.) did not hold; that, indeed, σ_{21} could be smaller than both σ_{01} and σ_{02}, or, in other words, that the judgments of independent observers could be sensibly correlated! I accordingly felt compelled to discard the current theory entirely, and develop one in which the correlations like r_{12}, &c., are not supposed to be zero. Before describing this, however, I must point out that even if, on the ordinary view, we put these correlations zero, we ought to expect correlation in the judgments of observers when they are referred to the judgment of a standard observer.

This may be proved thus:—

Let $\eta_1 = p_{01} - (x_1 - \xi)$, with similar values for η_2 and η_3. Then $S(\eta_1) = S(\eta_2) = S(\eta_3) = 0$; $S(\eta_3^2) = n\sigma_{03}^2$; $S(\eta_2\,\eta_3) = n\sigma_{02}\,\sigma_{03}\,r_{23} = 0$, since $r_{23} = 0$, and similarly $S(\eta_3\,\eta_1)$ and $S(\eta_1\,\eta_2) = 0$.

From (iii.) we have:

$$\rho_{3,\,12} = \frac{S\{(p_{31} + p_{01} - p_{03} + \eta_3 - \eta_1)(p_{32} + p_{02} - p_{03} + \eta_3 - \eta_2)\}}{n\sigma_{31}\sigma_{32}} = \frac{S(\eta_3^2)}{n\sigma_{31}\sigma_{32}},$$

remembering that $p_{31} = p_{03} - p_{01}$, $p_{32} = p_{03} - p_{02}$, and the relations cited above.

Hence:
Similarly:†

$$\left.\begin{aligned} \rho_{3,\,12} &= \sigma_{03}^2/(\sigma_{31}\sigma_{32}), \\ \rho_{2,\,31} &= \sigma_{02}^2/(\sigma_{21}\sigma_{23}), \\ \rho_{1,\,23} &= \sigma_{01}^2/(\sigma_{12}\sigma_{13}) \end{aligned}\right\} \quad \text{(viii.)}.$$

These expressions can never vanish, and thus, if the current theory were true, the judgments of two observers referred to a third as standard would undoubtedly be

* As, for example, by an artificial star, whose actual position at each instant of time is known, first, I think, used by N. C. Wolff in 1865. Unfortunately the personal equation seems to vary a good deal with the speed and intensity of the star observed.

† Of course relations of the type $\sigma_{13}^2 = \sigma_{03}^2 + \sigma_{01}^2$ will also hold by (vi.) if there be no correlation of absolute judgments.

correlated. Independence of absolute judgments connotes correlation of relative judgments. This is, of course, an instance of what I have termed "spurious" correlation, but it is none the less important that it should not be overlooked. When we cannot form absolute judgments, but refer our observations to a special observer as standard, then the observations so reduced of two independent observers will certainly be correlated. I am not aware that attention has hitherto been paid to this point when the observations of different observers relative to a standard man have been combined.

One result of the actual correlation of independent judgments is that the values experimentally determined for the ρ's are not those given by (viii.). A genuine correlation is superposed on the spurious correlation, and the total correlation observed may be greater or less than the values indicated in (viii.).

(4.) *New Theory of Errors of Observation.*

Let us suppose that the correlations r_{32}, r_{13}, r_{21} are not zero, then, provided we calculate the standard deviations of the absolute and relative judgments, we can find at once these correlations. We have

$$\left.\begin{aligned} r_{23} &= \frac{\sigma_{02}^2 + \sigma_{03}^2 - \sigma_{23}^2}{2\sigma_{02}\sigma_{03}}, \\ r_{31} &= \frac{\sigma_{03}^2 + \sigma_{01}^2 - \sigma_{31}^2}{2\sigma_{03}\sigma_{01}}, \\ r_{12} &= \frac{\sigma_{01}^2 + \sigma_{02}^2 - \sigma_{12}^2}{2\sigma_{10}\sigma_{02}} \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (\text{ix.}).$$

We are no longer able to find the absolute variabilities from the relative variabilities, and we require direct experiments in which the errors of absolute judgment are known in order to determine the correlations.

Turning now to the correlations between relative judgments, we easily deduce from (iii.)

$$\rho_{3,12} = \frac{\sigma_{03}^2 + r_{12}\sigma_{01}\sigma_{02} - r_{31}\sigma_{03}\sigma_{01} - r_{32}\sigma_{03}\sigma_{02}}{\sqrt{(\sigma_{03}^2 + \sigma_{01}^2 - 2\sigma_{03}\sigma_{01}r_{31})}\sqrt{(\sigma_{03}^2 + \sigma_{02}^2 - 2\sigma_{03}\sigma_{02}r_{32})}}$$

$$= \frac{\sigma_{31}^2 + \sigma_{32}^2 - \sigma_{12}^2}{2\sigma_{31}\sigma_{32}},$$

since $$\sigma_{31}^2 = \sigma_{03}^2 + \sigma_{01}^2 - 2\sigma_{03}\sigma_{01}r_{31},$$

and similar relations hold.

We have thus the series:

$$\left.\begin{aligned}\rho_{1,23} &= \frac{\sigma_{12}^2 + \sigma_{13}^2 - \sigma_{23}^2}{2\sigma_{12}\sigma_{13}} \\ \rho_{2,31} &= \frac{\sigma_{23}^2 + \sigma_{21}^2 - \sigma_{31}^2}{2\sigma_{23}\sigma_{21}} \\ \rho_{3,12} &= \frac{\sigma_{31}^2 + \sigma_{32}^2 - \sigma_{12}^2}{2\sigma_{31}\sigma_{32}}\end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (\text{x.}).$$

These suffice to find the ρ's as soon as a series of experiments giving relative judgments has been carried out. They will not suffice to differentiate the real and the spurious parts of the correlation between the relative judgments.

These results are, of course, quite independent of any theory of normal distribution. The correlation coefficients will give the probable value of an error of judgment which A will make when we know the error that B has made in the same observation. Thus, if e_{02} be the average error made by a second observer when a first makes the error e_{01}, we shall not have e_{02} equal to the personal equation of the second observer, but given by

$$e_{02} = p_{02} - p_{01}r_{12}\frac{\sigma_{02}}{\sigma_{01}} + e_{01}r_{12}\frac{\sigma_{02}}{\sigma_{01}} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (\text{xi.}).$$

Again, if e_{12} be the average error made by a second observer relative to a first, when a third observer makes an error relative to the first of e_{13}, then e_{12} will not be equal to the relative personal equation of the second observer, but must be determined from

$$e_{12} = p_{12} - p_{13}\,\rho_{1,23}\frac{\sigma_{12}}{\sigma_{13}} + e_{13}\,\rho_{1,23}\frac{\sigma_{12}}{\sigma_{13}} \quad . \quad . \quad . \quad . \quad . \quad . \quad (\text{xi. } bis).$$

It will thus be clear that the reduction of isolated observations to a common standard depends essentially on a discovery of the intensity of correlation for absolute or relative errors. $e_{02} = p_{02}$ will only be true when judgments have been shown to be perfectly independent. $e_{12} = p_{12}$ will practically be never true, for the ρ's can only vanish in the exceptional case in which the spurious and real correlations just balance each other's influence.

We shall find as we advance need to develop this theory in certain directions, but its main features have now been sufficiently indicated, and we can turn to the experimental results.

(5.) *General Description of the Experiments.*

The first series of experiments were made in the summer of 1896 by Dr. ALICE LEE, Mr. G. U. YULE, and myself. They were very simple in character. Sheets of white paper ruled with faint blue lines were taken, such as are sold for "scribbling," and on each blue line two segments of a line were obtained by pricking with a needle point. This was done in triplicate by running the needle point through three adjusted sheets. These segments formed a random distribution of lengths placed on a series of horizontal lines. Each observer now took 500 such lines—the series being the same for each—struck a pencil stroke with a fine pencil through the needle points terminating each segment, and then bisected that segment with a third pencil stroke at sight. We thus obtained three series of estimates of the midpoints of the same group of lines by three apparently independent observers. The judgments were made in the same room, under practically the same conditions of light for each individual, but each experimenter was not necessarily bisecting the same line at the same instant of time. The common factors were the length of the line and its position relative to the edge of the paper, which latter varied from line to line. It does not appear to me that these factors are more or less influential than the sameness of influences which must ever arise when two or more individuals judge the same phenomenon.

The actual length of the lines and the distance from the left-hand terminal of the point guessed as midpoint were now very carefully measured ; whatever errors occur in these measurements, and of course such must exist, they are of a totally different order of magnitude to the errors of midpoint judgment.* The letter u will be used to denote the length of any line, x for the distance from the left-hand terminal to the experimental bisection, $x' = x - \frac{1}{2}u$ will stand for the error in placing the midpoint, considered positive when towards the right. The subscript 1 refers to Dr. LEE's judgment, the subscript 2 to my judgment, and the subscript 3 to Mr. YULE's judgment. I should have liked to have taken 1000 instead of 500 judgments, but the labour of experimenting, and especially also of arithmetical reduction is so great that we had to limit ourselves to the smaller number. Even that, I believe, is far greater than has yet been used in the determination of personal equation.

À priori, it seemed reasonable to me that the longer the line the greater would be the error of its bisection. Accordingly x'/u, or the ratio of the error to the length of the line, was taken in the first place as the quantity to be tabulated. I call this quantity X′. Dr. LEE spent several months of the summer of 1896 in the reduction of the observations on this basis, and the series of diagrams giving the frequency curves were drawn for X′. The reduction, however, showed at once that the values of X′ for different observers were correlated. Such correlation of what I then thought must be independent judgments led me to more closely investigate the matter. I attributed this correlation of independent judgments to spurious correlation due to the use of indices, and I determined to reconsider the subject on an entirely different experimental plan, after developing the theory of spurious correlation.†

* That judgment was made rapidly as soon as the needle points terminating the line had been marked so as to be visible.

† See 'Roy. Soc. Proc.,' vol. 60, p. 489.

With the aid of Mr. HORACE DARWIN I arranged a series of experiments which should test simultaneously the eye, the ear, and the hand, and thus give every opportunity for a variety of small causes to influence the errors of judgment. My plan was as follows: A beam of light of very small breadth should traverse a white strip and at some part of its course a bell should sound. At this instant the eye should judge its position on the strip and the observer should at once divide a similar strip by a pencil stroke into parts in the same ratio as he considered the beam to divide the first strip. The instant at which the bell would sound was unknown to the observers, but it was so arranged that the exact position of the beam when the bell sounded could be easily ascertained by another person.

Mr. DARWIN constructed for us a pendulum,* consisting of a bar swinging on knife edges from an axis through its middle point. At either end of the bar were weights, so that by their adjustment very slow or very quick swings could be obtained. The pendulum could be released from rest at any angle from the vertical. Attached to the bottom of the pendulum was a small bell, which struck a very light hammer as it passed through the lowest point of the swing. This hammer was easily adjustable and was pulled upright by a string between each experiment, being knocked over by the transit of the pendulum. A mirror swinging about a horizontal axis had a strut attached to this axis and perpendicular to the plane of the mirror. This strut rested on a saddle (a) attached to a similar strut perpendicular to the pendulum bar at its axis. By shifting the saddle on the strut the mirror could be made to swing through a very small or a fairly large angle, whatever might be the amplitude of the pendulum. The whole object of this arrangement was to obtain a great variety of speeds and ranges for the line of light on the strip and so ascertain how far these conditions interfered with the independence of judgment which, *à priori*, I supposed must exist. When the first series of experiments showed substantial correlation in judgment, although the bright line moved in the same manner, no further series were then undertaken to determine how this correlation would be varied by differences of speed and range. Correlation existed when all the circumstances were alike except the position of the bright line on the strip when the bell sounded. I believed that I had evidence that the source of the correlation was rather in the observer than in the likeness of condition for each observer in each individual experiment,† and this was too subtle to be analysed by simply varying speed and range.

A beam of light from an electric lantern was intercepted by a screen having a thin horizontal slit placed in the slide groove; the selected part of the beam reflected from the pendulum mirror was received on a black screen at some distance from the

* See figure 1, p. 249.

† I hope later to take a further series of estimates, but it must be remembered that 500 experiments are the least we can make for our present purpose, and that with varying conditions the labour of making them will be greater, while the exhausting work of reduction will not be lessened.

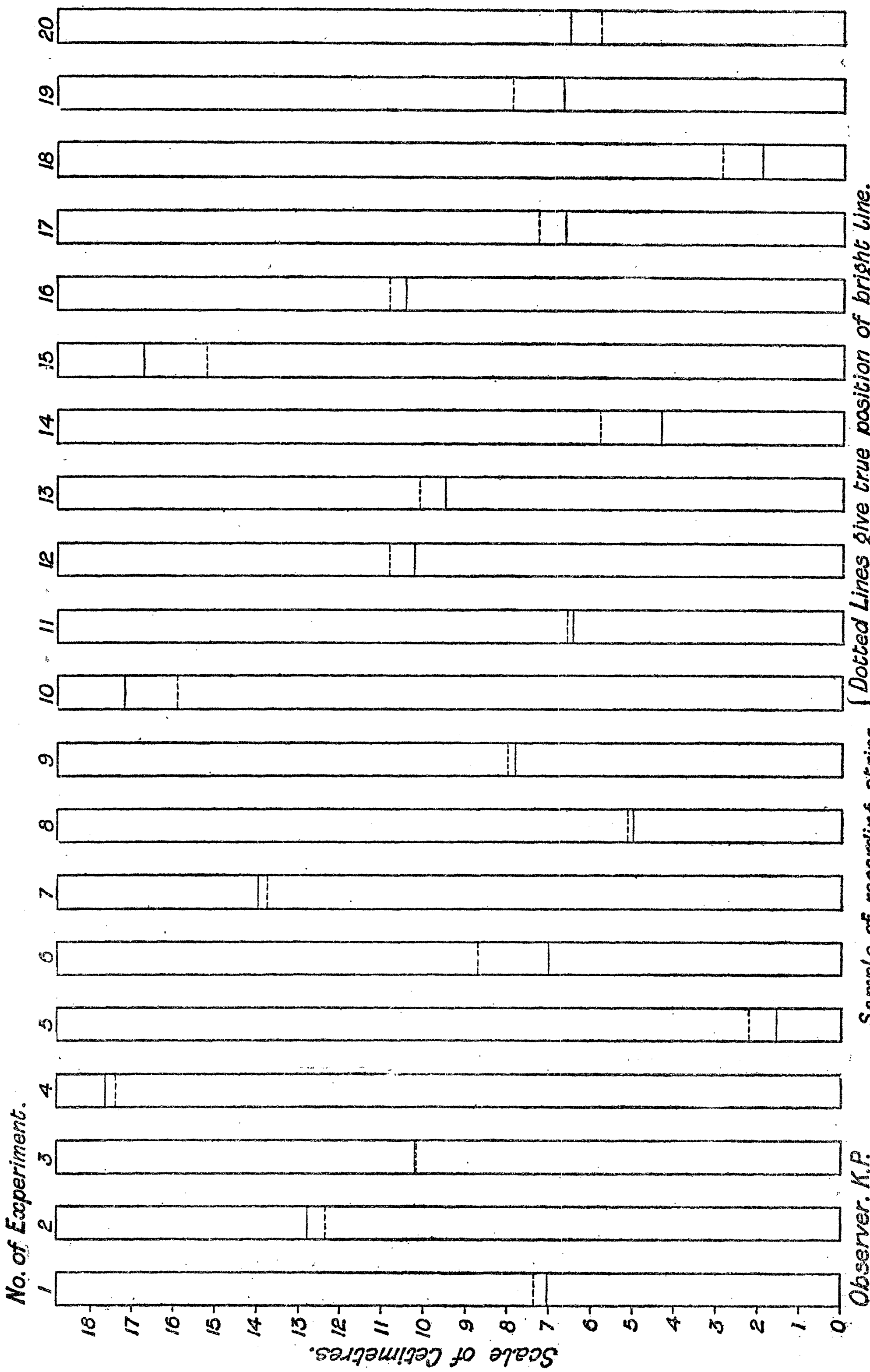
observers, and was practically absorbed, being invisible until a white strip was placed on the black screen; then the bright line was visible in a half-darkened room so long as it fell on the white strip. This white strip was 32·6 centims. long and 6 centims. broad; it could be placed anywhere on a scale painted in red on the black screen, and quite invisible to the observers.

The method of experimenting was as follows: The pendulum was brought to rest in a vertical position and the hammer was then moved up so as to touch the bell without resting against it, *i.e.*, it did not change its position when the pendulum was withdrawn. The line of light from the lantern now reflected from the stationary mirror fell on the scale on the black screen, which was adjusted by a fourth person so as to give a definite equilibrium position. The pendulum was now drawn back and clamped at a definite angle, which gave a very considerable range to the line of light. The three observers looking at the screen now saw no light at all, only 6 feet by 2 of black cloth. The fourth person now attached the strip of white card to the black cloth by aid of a drawing pin, so that its top coincided with any division on the scale known to himself only. He was thus able to make a record of the position on the strip occupied by the bright line when the hammer struck the bell. No doubt slight errors of adjustment occurred, but they were of much higher order than the errors of judgment. The equilibrium position of the beam was tested at the end of every twenty experiments, as well as the proper contact of the hammer.

A series of positions for the bright line on the strip were selected so as to cover fairly well the possible range, but the order in which these were taken was quite unknown to the observers. Of course, if the bell rung when the bright line just appeared on the strip, the latter was not moving as fast as if it rung when the bright line was just leaving the strip; but the range of the bright line was very considerable compared with the length of the strip, and I doubt whether this difference of speed was sufficient to sensibly influence the judgment.* The shifting of the strip on the screen was only adopted after it had been found that to adjust the equilibrium position of the bright line between each experiment to a fresh position on the screen-scale would mean an expenditure of time which it was impossible to provide for. It was easy enough to shift the equilibrium position, but it required two persons, one at the pendulum and one at the screen, to adjust the equilibrium position to a definite point of the scale, and the one at the screen instructing the other at the pendulum how to raise or lower the line of light in adjustment was likely, besides the evil of tediousness, to have far more influence upon the judgment of the observers than the fairly small shift of the strip while it was hidden from sight by the body of the adjuster.

Each observer was provided with a white sheet of paper on which were twenty

* If the correlations of judgments had been solely due to an "external cause" such as this, then it would not have been possible for the correlation to have been sensibly zero between two observers, but finite between the third observer and each of them.



Sample of recording strips. {Dotted Lines give true position of bright line. / Continuous Lines give estimated position of ditto.}

The scale on the left was added after the observations had been recorded.

rectangles similar to the white strip on the black screen, and he drew across these "recording" strips a line in the position he considered the line of light to have on the observation strip when the bell sounded. Every strip already used was covered up before a new observation was made, so that it might not influence the next judgment; the lines were drawn from left to right and all measurements taken on the left-hand side of the strip. A facsimile of one of the sheets of observations accompanies this paper, and will give graphically an idea of the nature of the errors of judgment made. These errors were then scaled off to the nearest tenth of a millimetre, and formed the basis of the second series of errors of judgment. The line of light travelled *down* the strip, and if the estimated line is below the real line on the recording strip the error was considered positive. If the personal equation were solely due to reaction time, this positive error would represent a lag of the judgment, *i.e.*, the bright line would be recorded as occupying a position posterior to what it really occupied when the bell sounded. A glance at the observations, however, shows that reaction time must have had very small influence on the total magnitude of the personal equation; two observers made rather large negative mean errors, and the third only a very small positive mean error.

The experiments were carried out in about a week, not more than 2 hours being given to them at a time, to prevent over-fatigue. The observers were Dr. ALICE LEE, Dr. W. R. MACDONELL, and myself. Mr. K. TRESSLER kindly acted as adjuster of the scale. The observers were screened from each other, but the experiments being conducted in a long narrow room, the only one available, Dr. LEE was placed somewhat further from the observation strip than Dr. MACDONELL or myself. The only other differentiation between the observers, that I am aware of, was that I released the pendulum from its clamp with my left hand, drawing the recording line with my right; the bright line moved so slowly, however, that I was not at all conscious of being hurried, and, as a rule, I had my left hand on the table before the line of light had entered the strip.

As the arrangement of the pendulum seems likely to be of service for similar observations, especially in the psychological laboratory, it is figured on the opposite page.

In this series of experiments, which will be termed the "bright-line series" to distinguish it from the "bisection series," x represents the error of judgment considered positive as defined above, and the subscripts 1, 2, 3 refer respectively to me, Dr. MACDONELL, and Dr. LEE. Before entering into the details of these series, I shall consider some points bearing on the method of reducing material of this kind.

(6.) *On the Means and Standard Deviations of Grouped and Ungrouped Observations.*

It is well known that if the distribution of errors follows the normal law, the "best" method of finding the mean is to add up all the errors and divide by their number, the "best" method of finding the square of the standard deviation is to

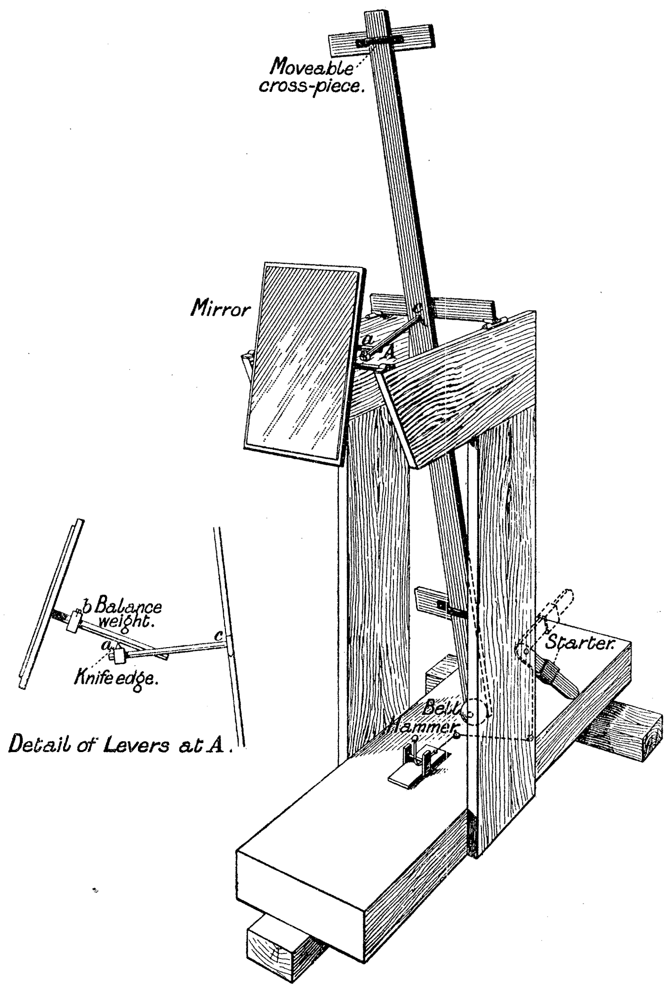


Fig. 1.—Apparatus for Personal Equation.

form the sum of the squares of the deviations from the mean and divide by their number, and the "best" method of finding a coefficient of correlation is to take the product of corresponding deviations from the respective means and divide by the product of the two standard deviations and the number of observations. These "best" methods become far too laborious in practice when the deviations run into hundreds or even thousands. The deviations are then grouped together, each group containing all deviations falling within a certain small range of quantity, and the means, standard deviations, and correlations are deduced from these grouped observations. If the means, standard deviations, and correlations be calculated from the grouped frequencies, as if these frequencies were actually the frequency of deviations coinciding with the midpoints of the small ranges which serve for the basis of the grouping, we do not obtain the same values as in the case of the ungrouped observations. It becomes of some importance to determine what corrective terms ought to be applied to make the grouped and ungrouped results accord. This point has been considered by Mr. W. F. SHEPPARD,* who has shown that from the square of the standard deviation we ought to subtract $\frac{1}{12}$th of the square of the base element of grouping, but that the mean and product of the grouped deviations should be left uncorrected. Thus corrected the values of the constants of the distribution as found from the ungrouped and grouped deviations will nearly, but not of course absolutely, coincide. In particular while the personal equation relation

$$p_{21} = p_{02} - p_{01}$$

will be absolutely satisfied for the ungrouped material, it will generally not be satisfied exactly for the grouped results. A test, however, of the practical justification for grouping is that the divergencies between the two methods ought to be of the order of the probable errors of the results. If this be so, then we may safely group. The fact that my grouped observations did not satisfy the relation cited above, led me to think it worth while that a comparison should at any rate be once made between ungrouped and grouped results on a large series of actual errors of observation. At the same time it gave me a means of verifying the accuracy of our very long arithmetical reductions by an independent investigation. The ungrouped observations were dealt with in the case of nine series involving 500 or 519 observations each. The labour of squaring so many individual deviations each read to four figures was lessened by using BARLOW's Tables, and the series were added up by aid of an American Comptometer, which for some years past we have found of great aid in statistical investigations.

(a.) *Bisection of Line Series.*

In Table I. will be found a comparison of the ungrouped and grouped results so far as the means and S.D.'s are concerned for our first series. X′ has been defined as the ratio of the error made in bisection to the length of the line bisected.

Here mean X_1' denotes that Dr. LEE made an average error of about 12/1000 of the length of the line in bisecting it, and that this error was to the *right* of the true midpoint. Mr. YULE and I made average errors of 4 to 5/1000 of a line in bisecting

* 'London Math. Soc. Proc.,' vol. 29, pp. 368, 375.

TABLE I.

500 trials.	Absolute personal equation.			Relative personal equation.		
	X_1'.	X_2'.	X_3'.	$X_2' - X_3'$.	$X_3' - X_1'$.	$X_1' - X_2'$.
Mean, ungrouped	+ ·01235 ± ·00074	− ·00444 ± ·00093	− ·00469 ± ·00079	+ ·00026	− ·01704	+ ·01679
" grouped	+ ·01230 ± ·00075	− ·00495 ± ·00093	− ·00377 ± ·00080	− ·00123 ± ·00098	− ·01589 ± ·00103	+ ·01712 ± ·00106
S.D., ungrouped	·02464 ± ·00053	·03068 ± ·00065	·02618 ± ·00056			
" grouped	·02455 ± ·00053	·03065 ± ·00065	·02625 ± ·00056	·03236 ± ·00069	·03376 ± ·00073	·03519 ± ·00075

it, and our errors were both to the *left* of the midpoint.* All these absolute equations are seen to be considerable multiples of their probable errors, or are undoubtedly significant. While Dr. LEE's personal equation is, roughly, three times as large as Mr. YULE's or mine, she is steadier in her judgment, our relative steadiness being as $\frac{1}{25} : \frac{1}{31} : \frac{1}{26}$ nearly, or about as 40 : 32 : 38.

The absolute personal equations show that the probable errors of the means and of the standard deviations are for all practical purposes identical, whether they are calculated from the standard deviations of the ungrouped or grouped observations. From these probable errors we see that the differences between the ungrouped and grouped results are in all cases but two less than the probable error of the quantity; in one of these cases, however, the difference is only very slightly greater, and accordingly it is not of any practical importance. In the other case, Mr. YULE's personal equation is insignificantly larger than mine for ungrouped results, and slightly smaller than mine for grouped results. The effect of this is that our relative personal equation swings round from negative to positive as we pass from ungrouped to grouped deviations. The total change is only ·00149, and as the probable error of the result is ·00098, we are perhaps hardly justified in holding that the grouped results are in disagreement with the ungrouped. I think all we could say is that our absolute personal equations are very nearly equal, and that we have sensibly no relative personal equation. The differences of the other relative personal equations as found by the two methods are less than their probable errors.†

* The light fell from the left hand on the paper for all three experimenters during the bisections.

† The reader will notice at once that the relation $p_{23} = p_{02} - p_{03}$ no longer holds. If we deduce the relative from the absolute personal equations we find:

$$p_{23} = -\cdot 00118,\ p_{31} = -\cdot 01607 \text{ and } p_{12} = +\cdot 01725 \text{ instead of}$$
$$-\cdot 00123 \qquad -\cdot 01589 \text{ and } +\cdot 01712 \text{ respectively.}$$

The differences are, however, quite insignificant, when we consider the probable errors.

So far, then, as this first series of experiments goes, we have ample justification for grouping our deviations.

(*b*.) *Bright Line Series.*

In this case I compared the results for ungrouped and grouped observations not only as far as concerns absolute personal equations, but also for the relative personal equations, and even for the coefficients of correlation. We have therefore a still wider basis for drawing inferences. This is done in Table II., x being now the error, positive if the bright line is recorded on the strip as being below its true position.

To find a length on the observation strip from that on the recording strip we have to multiply by the factor 1·734.

TABLE II.

519 observations.	Absolute personal equation.			Relative personal equation.		
	x_1.	x_2.	x_3.	$x_2 - x_3$.	$x_3 - x_1$.	$x_1 - x_2$.
Mean, ungrouped	+ ·06724 ± ·03538	−1·14906 ± ·03480	− ·48563 ± ·05377	− ·66343 ± ·05170	− ·55287 ± ·05954	+1·21630 ± ·04928
,, grouped .	+ ·07774 ± ·03521	−1·14483 ± ·03473	− ·44635 ± ·05393	− ·68275 ± ·05148	− ·61145 ± ·05943	+1·21518 ± ·04932
S.D. ungrouped .	1·19495 ± ·02502	1·17546 ± ·02461	1·81599 ± ·03802	1·74616 ± ·03656	2·01091 ± ·04210	1·66454 ± ·03485
,, grouped .	1·18913 ± ·02489	1·17289 ± ·02455	1·82146 ± ·03813	1·73883 ± ·03640	2·00717 ± ·04202	1·66597 ± ·03488
Correlations.	r_{23}.	r_{31}.	r_{12}.	$\rho_{1,\,23}$.	$\rho_{2,\,31}$.	$\rho_{3,\,12}$.
Ungrouped . .	·3819 ± ·0253	·1571 ± ·0289	·0139 ± ·0296	·5625 ± ·0202	·3055 ± ·0268	·6154 ± ·0184
Grouped . . .	·3908 ± ·0251	·1624 ± ·0288	·0051 ± ·0296	·5653 ± ·0201	·3056 ± ·0268	·6127 ± ·0185

We see at once from this table that the probable errors of means, standard deviations, and correlations are for all practical purposes the same whether we group the observations or not. In the next place we find that, judged by these probable errors, the differences are less than would arise from the results of random sampling. Thus in all cases the differences are less than the probable errors, and in most cases very considerably less. The greatest divergence occurs in the relative personal equation of Dr. LEE and myself, but even in this case the difference is just less than the probable error. We may accordingly conclude that with such a number of groups as we are here using, we may safely group observations, and the differences between the constants calculated from the absolute formulæ and from the grouped results will not exceed such errors as must arise from our statistics being a random sample and not embracing the entire "population" of errors.

The interpretation of Table II., to which we shall frequently have occasion to refer, may be given here. Abiding by the ungrouped data and multiplying by 1·734, we find for the observation strip of 32·6 centims. the results:

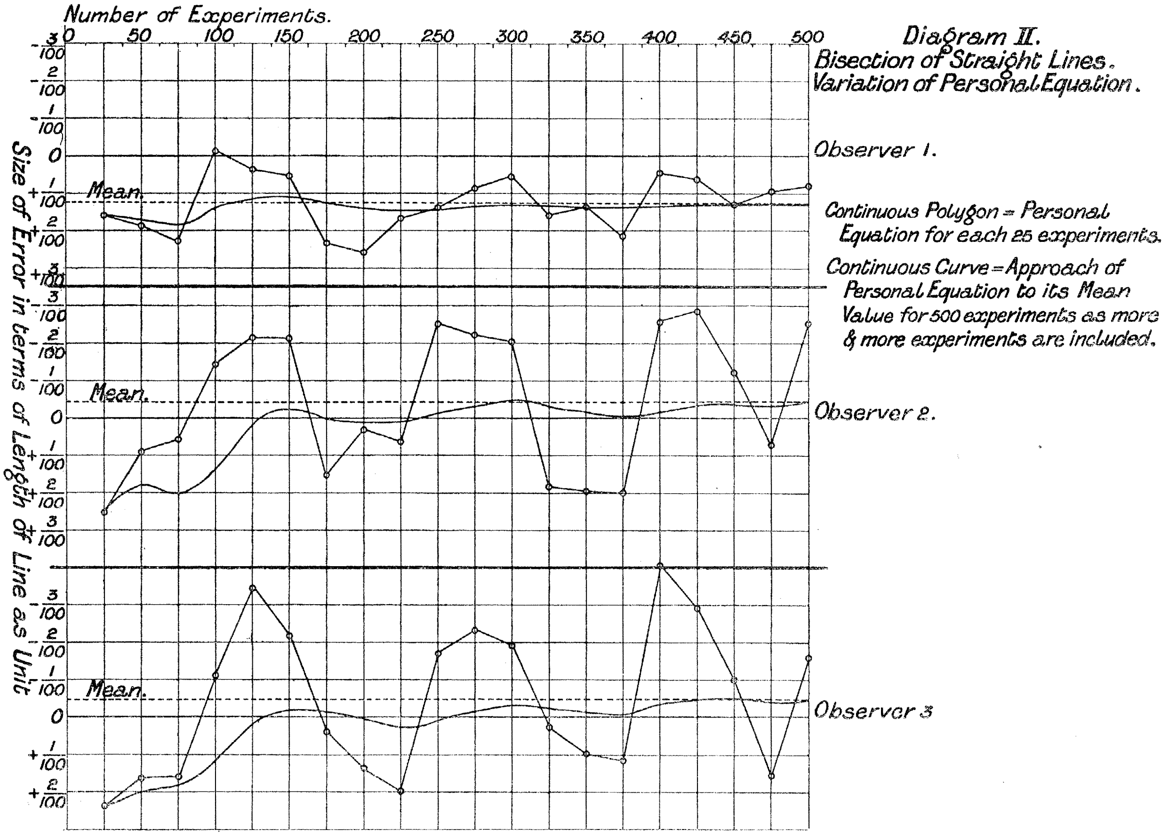
Observer.	Mean.	Standard deviation.
Professor PEARSON . . .	+ ·1348	2·0620
Dr. MACDONELL	−1·9852	2·0333
Dr. LEE	− ·7740	3·1585

Thus on an average I was 1·3 millims. ahead of the true position; such a personal equation might arise from a reaction time. On the other hand, Dr. MACDONELL anticipated the position of the ray by 19·8 millims. on the average, and Dr. LEE by 7·7 millims. Their personal equations cannot, therefore, be due to reaction time. Dr. MACDONELL is slightly steadier in his judgment than I am, and we are both considerably steadier than Dr. LEE. She and I have about changed our relative positions; her steadiness is to mine in the ratio of about 40 to 32 in bisecting lines, but as 26 to 40 in judging of the position of a bright line on a scale. This change of position with regard to steadiness may be due to the different nature of the two series of experiments, or to the lapse of time, 4–5 years, between the two. Dr. MACDONELL with the largest personal equation is the steadiest of the three observers in his judgment. It is noteworthy that in both sets of experiments the observer with the largest personal equation judges most steadily. So far as our results reach, there appears to be no marked relationship between accuracy and steadiness of judgment.

(7.) *On the Constancy of the Personal Equation.*

The totals of our results were for the ungrouped returns added up first for every twenty-five to fifty trials, and this enables us to appreciate the degree of constancy in the personal equation when it is determined as it actually is, and probably must be in practice, from a comparatively few experiments.

Table III. (p. 256) gives the changes in personal equation for the three observers as based upon every series of twenty-five bisections, and, further, the personal equation as based upon 25, 50, 75, 100, 125, . . . 475, 500 experiments. These results are represented graphically in Diagram 2. In this diagram 1 unit of the vertical scale represents an error of only $\frac{1}{100}$th of the length of the line in placing its midpoint. It will be noticed that if we take 200 experiments, the variation in the value of the personal equation obtained by taking any larger number scarcely amounts to $\frac{1}{200}$th of the length of the line bisected. On the other hand, it must be noted that the fluctuations in the personal equation when we come to deal with series of 25 are much larger than the probable error of a random sampling. The probable error of the personal equation of Dr. LEE, based on 25 experiments, is ± ·00331, but actually the personal equation as determined from two different sets of 25 experiments may amount to seven or eight times this amount. In other words, there is a significant difference in personal equation depending upon the individual 25 lines bisected



Whether this significant difference is due to the lengths of those lines, their exact position on the paper, or to the individual state of the observer, it may be hard to determine. It may even be due to slight variations in light occurring between one 25 series of experiments and the next. But whatever be the single source or combination of sources to which these changes of personal equation are due, it seems to me that they are so insignificant and subtle that they will occur in almost every kind of physical measurement we may take. It would be idle to attempt indeed to discover and eliminate such sources, for while it might be possible after elaborate investigation to eliminate them in an especially devised series of experiments, this could not be done in practice, where we must take our observer's experiments as they are given to us, and where we cannot possibly ensure uniformity in light, in mood, or health of observer, and as well as in all the features of the observed phenomena. The true conclusion appears to be that the range of data upon which the personal equation is based must be very wide, so as to swamp as far as possible these sources of variation due to the "atmosphere" of a short consecutive series. But if such a personal equation be found, what will be its value? It can hardly be applied satisfactorily to an isolated observation or to a short consecutive series of observations, for these will of course be influenced by their special atmosphere. It would only have value for a long series such as it was itself determined for, and such a series would rarely occur in practice. The fact that in both our series of experiments the differences between the values of the personal equation as found from short series are many times the probable error of sampling is very remarkable. I shall refer to it as "the influence of immediate atmosphere," where I understand the "atmosphere" to be compounded of all the little sources which affect either the observed thing or the observer more or less persistently during a short series. I am prepared to be told that the influence of immediate atmosphere was something peculiar to our own test experiments. But I shall require a good deal of the hard logic of experimental facts to be convinced that it has no existence in astronomical observations. There are many determinations of astronomical personal equation, but in published data I have been unable to discover enough material to determine how far the admitted variations in personal equation for short series are or are not of the order of deviations due to random sampling.

The following data will bring out the points of this discussion :—

TABLE V.—Personal Equation.

Observer.	First series.				Second series.			
	Experiments.	Bisection of lines.			Experiments.	Position of bright line.		
		1.	2.	3.		1.	2.	3.
Mean	500	+ ·01235	− ·00444	− ·00469	519	·06724	−1·14906	− ·48563
Ditto	1–250	+ 01424	− 00173	+ ·00065	1–266	·09571	−1·17282	− ·28940
Ditto	251–500	+ ·01046	− ·00714	− ·01004	267*–520	·03731	−1·12407	− ·69194
Standard deviation . .	500	·02464	·03068	·02618	519	1·19495	1·17546	1·81599
Probable error of mean	500	·00074	·00093	·00079	519	·03538	·03480	·05377
Ditto	250	·00105	·00132	·00112	260	·04998	·04917	·07596
Ditto	25	·00331	·00416	·00353	30	·14715	·14475	·22369

* Experiment 291 omitted.

This table shows us :—

(*a.*) That the probable errors for the personal equations deduced from twenty-five bisections are such that the fluctuations of personal equation given in Table III. or Diagram 2 are in very many cases significant.

TABLE III.—Personal Equation in Bisection of Lines.

Experiments.	Observer 1.		Observer 2.		Observer 3.	
	(a)	(b)	(a)	(b)	(a)	(b)
1–25	+ ·01548	+ ·01548	+ ·02627	+ ·02627	+ ·02379	+ ·02379
26–50	+ ·01828	+ ·01688	+ ·00890	+ ·01759	+ ·01600	+ ·01989
51–75	+ ·02252	+ ·01876	+ ·00596	+ ·02056	+ ·01585	+ ·01855
76–100	− ·00149	+ ·01370	− ·01429	+ ·01342	− ·01149	+ ·01104
101–125	+ ·00340	+ ·01164	− ·02163	+ ·00104	− ·03513	+ ·00180
126–150	+ ·00498	+ ·01053	− ·02112	− ·00265	− ·02196	− ·00216
151–175	+ ·02326	+ ·01235	+ ·01524	− ·00010	+ ·00381	− ·00130
176–200	+ ·02576	+ ·01402	+ ·00290	+ ·00028	+ ·01326	+ ·00052
201–225	+ ·01629	+ ·01428	+ ·00649	+ ·00097	+ ·01979	+ ·00266
226–250	+ ·01396	+ ·01424	− ·02604	− ·00173	− ·01739	+ ·00065
251–275	+ ·00822	+ ·01370	− ·02217	− ·00359	− ·02387	− ·00158
276–300	+ ·00514	+ ·01298	− ·02105	− ·00504	− ·01977	− ·00309
301–325	+ ·01595	+ ·01321	+ ·01814	− ·00316	+ ·00274	− ·00264
326–350	+ ·01323	+ ·01321	+ ·01943	− ·00164	+ ·00944	− ·00178
351–375	+ ·02140	+ ·01376	+ ·02003	− ·00020	+ ·01132	− ·00091
376–400	+ ·00484	+ ·01320	− ·02636	− ·00183	− ·04012	− ·00336
401–425	+ ·00614	+ ·01279	− ·02902	− ·00343	− ·02904	− ·00487
426–450	+ ·01320	+ ·01281	− ·01227	− ·00392	− ·00978	− ·00514
451–475	+ ·00946	+ ·01263	+ ·00073	− ·00333	+ ·01512	− ·00408
476–500	+ ·00698	+ ·01235	− ·02548	− ·00444	− ·01643	− ·00469

TABLE IV.—Personal Equation for Position of Bright Line.

Experiments.	Observer 1.		Observer 2.		Observer 3.	
	(a)	(b)	(a)	(b)	(a)	(b)
1–37	+ ·29973	+ ·29973	− ·27703	− ·27703	+ ·42243	+ ·42243
38–74	+ ·04838	+ ·17406	− ·50540	− ·44122	+ ·65919	+ ·54081
75–111	+ ·49432	+ ·28081	− 1·58568	− ·82270	+ ·22459	+ ·43541
112–148	+ ·13595	+ ·24459	− 1·41649	− ·97115	+ ·42162	+ ·43196
149–185	+ ·02459	+ ·20059	− 1·18027	− 1·01297	− 1·02459	+ ·14065
186–212	− ·03074	+ ·17113	− 1·61037	− 1·08906	− 1·15704	− ·02462
213–239	− ·09185	+ ·14142	− 1·60111	− 1·14690	− 1·73926	− ·21707
240–266	− ·30889	+ ·09571	− 1·40222	− 1·17282	− ·92963	− ·28940
*267–293	− ·40308	+ ·05130	− 1·21692	− 1·17675	− ·52346	− ·31024
294–320	− ·24519	+ ·02621	− 1·24593	− 1·18260	− ·67747	− ·34132
321–347	− ·54185	− ·01812	− ·97185	− 1·16616	− ·16556	− ·32760
348–374	+ ·08889	− ·01038	− 1·25704	− 1·17273	− ·66889	− ·35231
375–401	− ·04185	− ·01250	− ·92222	− 1·15582	− ·47518	− ·36060
402–435	− ·13735	− ·02228	− 1·07912	− 1·14982	− 1·22735	− ·42850
436–469	+ ·48117	+ ·01429	− ·86324	− 1·12900	− 1·34088	− ·49479
470–503	+ ·52029	+ ·04857	− 1·45265	− 1·15092	− ·39588	− ·48809
504–520	+ ·61882	+ ·06724	− 1·09412	− 1·14906	− ·41294	− ·48563

Columns (*a*) contain the personal equations determined from the experiments given in the first column. Columns (*b*) contain the personal equations determined from all the experiments up to and including the last given in the corresponding line of the first column.

* Not including No. 291, which was *à priori* rejected. Hence the total number of experiments dealt with is one less after this than the number recorded to the right of the first column.

Although the personal equations in Table IV. are based upon series varying in number from 26 to 37, the probable errors for thirty observations of the position of a bright line suffice to show that the fluctuations in the values of the personal equations as given in Table IV. or in Diagram III., p. 270, are in many cases significant.

(*b*.) The probable errors for the personal equation in bisecting 250 lines show that there were significant changes in the personal equations of the three observers between the first and second moiety of the experiments. While Dr. LEE (1) bettered her judgment by ·004, Mr. YULE (3) swung over from ·001 to right of true midpoint to ·010 to left of midpoint, and I had a worse judgment by ·005 in the second moiety when compared with the first moiety of the results.

In the case of the second series with the bright line, Dr. MACDONELL (2) and I (1) have changes slightly for the better in our judgments between the 266 first experiments and the 253 second experiments; but having regard to the probable errors given for 260 experiments, it may be doubted whether these changes are significant. Dr. LEE (3) has, however, a quite significant change for the worse.

The fact that in some cases the personal equation grows less, in others greater, in the second half of the series seems to indicate that the changes in personal equation were by no means due to a secular improvement in judgment.* Nor do they admit of explanation on the assumption of increasing fatigue due to the exhaustion of the power of attention. It must be remembered that the experiments were spread out over a number of days, and this cause would only influence the latter experiments on each day. My worst experiments on the bright line are the Series 321–347 and 504–520 (Observer (1) Column (*a*) Table IV.), but they are much above the average in goodness for Dr. LEE (Observer (3) Column (*a*)), and above the average for Dr. MACDONELL. Dr. MACDONELL'S worst results are 186 to 239 (Observer (2) Column (*a*)), and these, especially 213 to 239, are bad for Dr. LEE, but they are very good results so far as I am concerned. If any fluctuation was accordingly due to fatigue, it did not affect us alike.

While these fluctuations in short series are significant, they by no means screen the general features of each observer's individuality. Dr. LEE is clearly in the habit of bisecting straight lines at a point some $\frac{1}{100}$ or more to the right of the true point of bisection, while I place it with a sensibly less error to the left. She places a line of light moving downwards over a vertical strip ·8 centim. above its true position, and I about ·1 centim. below its true position at any instant. Dr. MACDONELL, on the other hand, with the steadiest judgment of all three, displaces it 2 centims. above its true position.† The differences of personal equation in both series for all three observers are quite significant when compared with the

* It should be noted that in the cases of Dr. LEE, Mr. YULE, and myself we have for years been accustomed to reading scales and judging proportional parts by the eye.

† Table V., second series, gives lengths on recording strip. The actual values for observing strip are given on p. 253.

probable errors of the differences, *i.e.*, there is a real individuality in observation which manifests itself in the personal equation.

But the fluctuations in the personal equation are significant too, and they cannot offhand be attributed to anything like betterment with practice, or decadence with fatigue.

So long as the variations in the constants of an experimental series can be shown to be within the errors of random sampling we feel on safe ground; we know the number of experiments required to obtain a result with any required degree of accuracy. On the other hand, when we find significant fluctuations in the personal equation depending on the influence of immediate atmosphere, it becomes all the more important to show in each individual investigation that the personal equation itself is insignificant. Let me illustrate this point. A physicist makes twenty or thirty measurements of a quantity, say by aid of a bright line moving across a scale. He gives the mean value m of the result and also what he terms its probable error e. Now the use of this probable error I take it to be this. If the same experiments were to be repeated by the same man the same number of times with the mean result m', then we should expect to find $m' - m$ not a large multiple of the probable error of the difference $\sqrt{(e^2 + e^2)} = \sqrt{(2)}\, e$. e gives us a test of the closeness with which the result will repeat itself on repetition of the experiments. But the whole foundation of this statement is the hypothesis that the twenty or thirty experiments dealt with are a *random sampling* of all possible experiments that might be made. Now the variability in the results of the individual experiments includes the variability of personal error, and the hypothesis supposes that the personal errors are a random sampling of the observer's personal errors. Our investigations seem to indicate that the personal errors are far from being a random sampling but depend in some subtle manner on the influence of immediate atmosphere. Hence, unless it can be shown that the latter influence is small as compared with other sources of error in the measurement under consideration, the mere calculation of the probable error is by no means a security for the same observer reaching the same result on repeating the original series of experiments.

We, of course, for both series selected experiments in which the personal error would be large,* and accordingly could be easily dealt with. But the division of scale lengths by the eye and the estimated position of a bright line are fundamental in many types of physical observation. Further, large errors are for theoretical purposes quite as good as, for practical purposes much better than, small, when we wish to obtain an answer to the question: Are the fluctuations in personal equation merely the result of random sampling, or are they due to the influence of immediate atmosphere?

So important is it to realise that these fluctuations are not due to random sampling,

* As a matter of fact only at a maximum $\frac{1}{100}$ in dividing a line and $\frac{1}{10}$ in determining the position of a bright line between two scale marks.

that I have worked out all the constants for the second series, for the whole set of experiments, and for its first and second moiety.

They are given in the accompanying table.

TABLE VI.—Influence on Constants of Fluctuations in Personal Equation.

		All Observations. 1–520 (without 291).			*First Series.* 1–266 inclusive.			*Second Series.* 267–520 (without 291).		
	Observer.	1	2	3	1	2	3	1	2	3
Absolute Judgments.	Mean	·0672	−1·1491	− ·4856	·0957	−1·1728	− ·2894	·0373	−1·1241	− ·6919
		± ·0354	± ·0348	± ·0537	± ·0514	± ·0520	± ·0778	± ·0484	± ·0459	± ·0728
	S.D.	1·1949	1·1755	1·8160	1·2428	1·2563	1·8815	1·1417	1·0833	1·7205
		± ·0250	± ·0246	± ·0380	± ·0363	± ·0367	± ·0550	± ·0342	± ·0325	± ·0516
	Correlation	·3819	·1571	·0139	3677	·2594	·0530	·4123	·0256	− ·0368
		± ·0253	± ·0289	± ·0296	± ·0355	± ·0386	± ·0412	± ·0352	± ·0424	± ·0423
	Observers.	3–2	1–3	2–1	3–2	1–3	2–1	3–2	1–3	2–1
Relative Judgments.	Mean	·6634	·5529	−1·2163	·8834	·3851	−1·2685	·4321	·7292	−1·1614
		± ·0517	± ·0595	± ·0493	± ·0760	± ·0814	± ·0711	± ·0683	± ·0865	± ·0680
	S.D.	1·7462	2·0109	1·6645	1·8385	1·9676	1·7198	1·6114	2·0406	1·6026
		± ·0366	± ·0421	± ·0348	± ·0538	± ·0575	± ·0503	± ·0483	± ·0612	± ·0481
	Correlation	·5625	·3055	·6154	·5085	·3900	·5935	·6323	·1938	·6375
		± ·0202	± ·0268	± ·0184	± ·0307	± ·0351	± ·0268	± ·0255	± ·0408	± ·0252

In the row in absolute judgments, entitled "Correlation," the correlation, r_{32}, of the judgment of the second and third observers is entered in column (1), r_{13} in column (2), r_{21} in column (3). In the row in relative judgments, entitled "Correlation," the correlation of the judgments of the second and third observers referred to the first observer as a standard, or $\rho_{1,\,23}$, is entered in column (1), $\rho_{2,\,31}$ in column (2), and $\rho_{3,\,12}$ in column (3).

The figures in antique type give the probable errors of each constant, and the probable errors in the differences of the constants can be found in the usual way as the square root of the sum of the squares of these.

Dealing first with the absolute observations, we note that the personal equations of Dr. MACDONELL and myself, (2) and (1), are within the limits of the probable errors the same for either half series and for the whole series. Both of us appear to have improved by about ·03, but whether this is a real improvement between the first and second series it is impossible to say, for the probable error of the half series is as much as ·05. In my own case, in the second series my personal equation is less than its probable error, and accordingly on the basis of 253 experiments—a number be it noted far larger than could ever be made in actual practice—it would be impossible to say whether I had a personal equation or not. I mention this point, because it seems to me a *sine quâ non* of all investigations of personal equation that the probable error of the results should be given, and in most cases one seeks for it in vain.

Dr. LEE's personal equation has increased substantially between the first and second series. All three observers have grown apparently steadier in their judgment. The probable errors, however, of the S.D.'s do not allow of the assertion that the steadiness has substantially increased. The variations in correlations of judgments are noteworthy. Judged from the first series, or the second series, or the whole series, the correlation between the judgments of Dr. LEE and Dr. MACDONELL remains sensibly the same, *i.e.*, ·4 within the limits of the probable error; there is sensibly no correlation between the judgments of Dr. MACDONELL and myself as given by any of the three series. Between Dr. LEE and myself there is on the whole series a substantial correlation of $\cdot 16 \pm \cdot 03$, but the two half series show us that it was on the wane during the course of the experiments, having fallen from the comparatively high value of ·26 to practically zero between the two half series. Whatever causes therefore produced the marked divergence of personal equation between Dr. MACDONELL and myself, they seemed to have been combined in Dr. LEE, and—to speak metaphorically—the dominant set for Dr. MACDONELL became after a struggle dominant for Dr. LEE; her methods of judging in the course of the experiments became more and more like Dr. MACDONELL'S and less like mine.

We turn now to the relative judgments. These it must be remembered are the only data which would be generally known in practice. Here it is only in the difference of Dr. MACDONELL'S and my judgments (column 2–1) that there is any real approach to constancy in the relative personal equation. The differences of our judgments have sensibly the same value for the first, the second, and the whole series. The same remark applies also to relative steadiness of judgment.* On the other hand, the relative personal equations of Dr. LEE and Dr. MACDONELL, or of Dr. LEE and me, differ substantially between the first half and the second half series. The relative steadinesses of judgment are less altered, being sensibly constant for Dr. LEE and myself, but possibly varying slightly for Dr. LEE and Dr. MACDONELL.

When we turn to the correlation of relative judgments, that of Dr. MACDONELL'S and my judgments, referred to Dr. LEE'S as standard, shows sensible constancy throughout the three series; that of Dr. MACDONELL'S and Dr. LEE'S, referred to mine as standard, shows not very large but sensible change; and finally that of Dr. LEE'S and mine referred to Dr. MACDONELL'S, shows very substantial modification.

Now judged by size of personal equation I stand first and Dr. MACDONELL last, judged by steadiness Dr. MACDONELL and I are almost equal (within the limits of the probable error), and Dr. LEE last. The most constant results for absolute personal equation are found—as we might *à priori* expect they would be—where the steadiness is greatest. But if we wish to obtain relative judgments whose relationship to each other will remain at closely the same value during a long series, then apparently we ought to refer not to the most steady, but to the least steady of the observers as a standard.

* I may remind the reader of what this exactly means: The differences of 1·72 and 1·60, the standard deviations for (2–1) in the first and second series, from 1·66, the standard deviation in the whole series, are about ·06, and this is just about the magnitude of the probable error of these differences, *i.e.*, $\sqrt{\{(\cdot 035)^2 + (\cdot 050)^2\}} = \cdot 061$.

It may be asked how, when as in practice we only know the relative judgments, are we to find out the degree of steadiness of the individual observers? This is a very important problem, and the answer would be perfectly clear if the old theory on p. 240 of this memoir were correct. Unfortunately the correlation of judgments comes in, and deprives us of any means of judging from a knowledge of relative steadinesses what the absolute steadinesses are. Let me illustrate this: The relative variabilities are greatest in the cases of 3–2 and 1–3, we might therefore suppose 3 to be least steady; the relative variabilities are least for 3–2 and 2–1, we might therefore suppose 2 to be most steady, and we should thus reach the actual scale of steadiness in absolute judgments—Dr. MACDONELL, myself, Dr. LEE. But now turn from the bright-line experiments in Table VI. to the bisection experiments in Table I. The relative judgment standard deviations are greatest for 3–1 and 1–2 and least for 2–3 and 3–1, we should therefore suppose that 1 was least steady and 3 most steady, or the order of steadiness 3, 2, 1, *i.e.*, Mr. YULE, myself, Dr. LEE. But an examination of the absolute standard deviations shows us that the real order is quite different, being Dr. LEE, Mr. YULE, and myself. In other words, no argument can be drawn, owing to the correlation in judgments, from relative to absolute steadiness.

It seems therefore impossible without experiments *ad hoc* to determine which observer is steadiest in judgment from a knowledge of relative personal equations.

We can only conclude that, at any rate in our own cases, the fluctuations in personal equation are such that, even in what are—for practical purposes—very large series, we cannot invariably assume them to be due to random sampling. We cannot attribute sensible changes in our own case to practice or to fatigue, but the high correlation of judgments suggests an "influence of the immediate atmosphere," which may work upon two observers for a time in the same manner.

(8.) *On the Interdependence of Judgments of the same Phenomenon.*

(i.) *The Bright-line Experiments.*

In the preceding paragraphs of this paper we have already had occasion to frequently refer to the correlation of the judgments of independent observers. Relations (vi.) of p. 240 are not fulfilled, nor even approximately fulfilled. For example, in Table II. we find $\sigma_{23} = 1 \cdot 74$, about, which is actually *less* than $\sigma_{03} = 1 \cdot 82$, about, when, if the theory of p. 240 were correct, $\sigma_{23} = \sqrt{(\sigma_{03}^2 + \sigma_{02}^2)}$! An examination of Table II. show us substantial correlations in two out of the three cases between absolute judgments. Now it is well to put somewhat more definitely what is meant by this correlation. Astronomers have already found that the brightness of a star influences the personal equation.* This in the language of the

* 'Monthly Notices of the Roy. Astron. Soc.,' vol. 60, November, 1899.

present writer produces a correlation of judgments, for "every one of the observers records the time of transit of faint stars later than that of bright stars." Hence if a number of observations were made on stars of varying magnitude, the judgment being a function of the magnitude, we should have a series of correlated errors. Again it is quite possible that the rate of transit of a bright line in our experiments might tend to correlate judgments, although the Cape observers did not find the personal equation to vary with stars of very different declination. It is not, however, contended that the correlation of jugdments is not due to one cause or another. The point of the present writer is this, that when every effort is made to eliminate large causes, such as varying brightness or rate of motion of the line in our own experiments, there still remains a multitude of small causes which produce correlation. It might be possible in an *ideal* series still further to eliminate some of these, but in practical observation we have to take a given phenomenon as it is, and we cannot possibly subtract from it the whole of its characteristic atmosphere. The next point to be noticed is, that whatever be these lesser causes of the characteristic atmosphere, *e.g.*, possibility of judging better the position of a bright line when it is nearer to one or another part of its range of visibility, or of bisecting a line of one length better than of another length—*they affect different observers in quite different manners.* Unlike the brightness of stars, the fluctuations of personal equation due to these causes are in themselves personal. Dr. MACDONELL and I have within the limits of error no correlation in our judgments of the position of a bright line. Dr. LEE and Dr. MACDONELL have a correlation as high as that of a measure made on a pair of brothers. In other words, correlation of judgments is a personal matter, just as personal equation itself. We could no doubt increase it by introducing variety in the observed phenomena—degree of brightness, degrees of speed—but beyond such causes capable of differentiation, there appear to be others, which I have classed as the influence of the immediate atmosphere, and which appeal to different personalities in different ways, and where there is a resemblance between certain features of two personalities produce correlation in their judgments. For example, A and B are alike in their sight, being slightly short-sighted we will say, B and C are alike in their nervous temperament, being able to judge more correctly if the bell rings after the bright line has been visible a rather longer time. There is thus an element of personality the same in A and B and another the same in B and C. The result would be that A's and B's judgments would be correlated, and also B's and C's judgments would be correlated, but not necessarily A's and C's. Something like this probably is what actually occurs in the case of Dr. MACDONELL, Dr. LEE, and myself. But it would be practically hopeless to try and discover the common elements in our personalities, and what in the immediate atmosphere of the experiments affected such elements. Even if, in a long and laborious series of experiments and reductions, we could discover the subtle causes of our correlations or non-correlations, the results would be of small value, for they would be personal to ourselves, and in actual observation they could not be eliminated from our own future experiments, nor could the like causes be determined for other observers. We are forced to admit, I think, that correlation is a personal character of every pair of observers, and to look upon it as a personal constant to be determined by experiment.

Here again, however, arises the very same point as we have considered in discussing absolute steadiness of judgment—we do not in practice know the absolute judgments, and so cannot find the correlation of absolute judgments, r_{23}, r_{31}, r_{12}. All we can do is to refer the judgments of two observers to a third as standard, and then measure the correlation of relative judgment. In this case we have a result which is not purely personal; we have superposed on the correlation due to a common element in personality, an element of "spurious correlation."

Taking the bright-line experimental result from Table VI., we have for 519 observations:—

$$r_{32} = \cdot 3819 \pm \cdot 0253, \qquad \rho_{1,\,32} = \cdot 5625 \pm \cdot 0202,$$
$$r_{13} = \cdot 1571 \pm \cdot 0289, \qquad \rho_{2,\,13} = \cdot 3055 \pm \cdot 0268,$$
$$r_{21} = \cdot 0139 \pm \cdot 0296, \qquad \rho_{3,\,21} = \cdot 6154 \pm \cdot 0184.$$

The latter series, all that we should usually know, enables us to form no opinion at all about the former. The absolute judgments of Dr. MACDONELL and myself have sensibly no correlation; our relative judgments have the greatest correlation of all—such are the masking effects of spurious correlation when judgments are referred to a third observer as standard!

If, from the values of the standard deviations of the absolute judgments, we calculate what would be the spurious correlations on the assumption that the absolute judgments are not correlated, we have by the method of p. 241:—

$$\bar{\rho}_{1,\,32} = \frac{\sigma_{01}^2}{\sqrt{(\sigma_{01}^2 + \sigma_{02}^2)}\sqrt{(\sigma_{01}^2 + \sigma_{03}^2)}} = \cdot 3918,$$

$$\bar{\rho}_{2,\,13} = \frac{\sigma_{02}^2}{\sqrt{(\sigma_{02}^2 + \sigma_{03}^2)}\sqrt{(\sigma_{02}^2 + \sigma_{01}^2)}} = \cdot 3811,$$

$$\bar{\rho}_{3,\,21} = \frac{\sigma_{03}^2}{\sqrt{(\sigma_{03}^2 + \sigma_{01}^2)}\sqrt{(\sigma_{03}^2 + \sigma_{02}^2)}} = \cdot 7013.$$

Hence $\rho_{1,\,32} - \bar{\rho}_{1,\,32} = \cdot 1707$, $\rho_{2,\,13} - \bar{\rho}_{2,\,13} = -\cdot 0756$, and $\rho_{3,\,21} - \bar{\rho}_{3,\,21} = -\cdot 0859$, or the effect of the correlation of absolute judgments is to increase in one case and decrease in the other two the spurious correlation. Without direct experiments *ad hoc* I see no way of determining from the usual data of the personal equation how much of the observed correlation of judgments may be due to a common element in the personality, and how much is really spurious. The two causes sometimes work in the same, sometimes in opposite directions.

The bright-line experiments show in a perfectly direct and simple manner that the correlation of absolute judgments is not wholly due to some external source influencing all observers in the same way, but is the result of a common element in the personalities of two observers. They further demonstrate the extreme difficulty in actual observations of separating without experiments *ad hoc* this psychological from the spurious correlation. These are precisely the points they were designed to elucidate.

(ii.) *The Bisection Experiments.*

I have indicated that it was the correlation in judgments of independent observers in the case of bisection that led to the second series or bright-line experiments. After these had demonstrated that the correlation of judgments was not wholly spurious correlation, it seemed desirable to reconsider the bisection experiments with a view to analysing more fully the character of the correlation exhibited by them.

The reader will remember that the error in judgment in their case was taken to be the displacement to the right of the true midpoint measured as a fraction of the total length of the line bisected. The following are the values of the correlations between the absolute judgments and between the relative judgments thus measured :—

TABLE VII.

$r_{23} = \cdot 3627 \pm \cdot 0262$	$\rho_{1,\,23} = \cdot 5615 \pm \cdot 0207$
$r_{31} = \cdot 1139 \pm \cdot 0298$	$\rho_{2,\,31} = \cdot 4980 \pm \cdot 0227$
$r_{12} = \cdot 2053 \pm \cdot 0289$	$\rho_{3,\,12} = \cdot 4379 \pm \cdot 0244$

Thus in every case the correlation has a quite sensible value.

I have pointed out that the absolute displacement of the midpoint by the experimenter was divided originally by the length of the line, because *à priori* we supposed that errors of bisection would be proportional to the length of the line bisected. But that when I had more fully realised the meaning of spurious correlation I saw that the whole of the above correlations might be really spurious in character, for they were the correlations of ratios having the same denominator. The experiments were accordingly put on one side until the bright-line experiments were concluded. It then seemed desirable to determine the correlations between the absolute displacements of the midpoints, and to find the magnitude of the correlation between the lengths of the lines experimented on and the errors made in their bisection. The labour of reducing again all the data would be excessive, and a very little consideration showed me that it was really unnecessary, if we knew the variation and distribution of the lengths of the lines bisected. Let u stand for the length of any one of the bisected lines, which as we have seen were a random sample. Then we have the following distribution :—

TABLE VIII.—Distribution of the 500 Experimental Lines.

Magnitude in $\frac{1}{2}$ inches.	Frequency.	Magnitude in $\frac{1}{2}$ inches.	Frequency.
3·00	5	5·75	44
3·25	7	6·00	37
3·50	4	6·25	22
3·75	12	6·50	18
4·00	20	6·75	13
4·25	28	7·00	9
4·50	30	7·25	13
4·75	48	7·50	0
5·00	58	7·75	3
5·25	64	8·00	4
5·50	58	8·25	3

Here the frequency corresponding to any magnitude m in half-inches denotes all the lines whose lengths fall between $m - \cdot 125$ and $m + \cdot 125$ half-inches. The lengths of the experimental lines had before this grouping been read off to the nearest $\frac{1}{200}$ of an inch.

From this frequency we found :—

$$m_u = \text{mean value of } u = 5\cdot 3165 \text{ half-inches.}$$

$$\sigma_u = \text{standard deviation of } u = \cdot 9513 \text{ half-inch.}$$

$$v_u = \sigma_u / m_u = \cdot 1789.$$

Now let x'_q = distance of experimental point of bisection from real midpoint of line, positive if it fall to the right, and x_q = distance from left-hand terminal of line to experimental point of bisection in the case of the qth observer. Let us write $X'_q = x'_q/u$, then X'_q is the ratio error which we had previously dealt with, and $X_q = x_q/u$.

Clearly $$x'_q = x_q - \tfrac{1}{2}u,$$

and if m_z denote the mean value of a variant z, we at once find :

$$\left.\begin{aligned} m_{X_q} &= m_{X'_q} + \cdot 5 \\ \sigma_{X_q} &= \sigma_{X'_q} \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(xii.)}$$

Now $$X'_q = X_q - \cdot 5 = x_q/u - \cdot 5.$$

Treating in the usual way variations as differentials, whose squares and products may be neglected, we have :—

$$\delta X'_q = \frac{\delta x_q}{m_u} - \frac{m_{x_q}}{m^2_u}\,\delta u \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(xiii.)},$$

whence squaring, summing for all possible values, and remembering the definitions of standard deviation and correlation coefficient we find :

$$\sigma^2_{X'_q}/(m_{X'_q} + \cdot 5)^2 = v^2_{x_q} + v^2_u - 2v_{x_q}v_u r_{ux_q} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(xiv.)},$$

where v_{x_q} stands for σ_{x_q}/m_{x_q}.

But the left-hand side of this equation is known from the previous reductions, $\sigma_{X'_q}$, $m_{X'_q}$ having the values in Table I.; v_u has just been determined. Hence a knowledge of v_{x_q} would enable us to find r_{ux_q} without the labour of further correlation tables. The values of x_1, x_2, x_3 had of course been measured in order to find x'_1, x'_2, and x'_3, so that all we required were their frequency distributions. They were as follows :—

TABLE IX.—Table of Frequencies of x_q.

Magnitude in $\frac{1}{2}$ inches.	Observer 1.	Observer 2.	Observer 3.	Magnitude in $\frac{1}{2}$ inches.	Observer 1.	Observer 2.	Observer 3.
1·35	—	1	1	3·00	52	39	31
1·50	5	8	6	3·15	31	28	30
1·65	8	5	6	3·30	20	23	20
1·80	6	15	12	3·45	21	17	16
1·95	16	29	21	3·60	12	10	10
2·10	29	36	39	3·75	10	6	10
2·25	42	47	49	3·90	3	4	3
2·40	41	54	60	4·05	2	2	2
2·55	56	62	63	4·20	3	1	2
2·70	75	55	56	4·35	—	1	1
2·85	68	55	62	4·50	—	2	—

Here the unit of grouping is ·15 half-inch, and a magnitude m covers all the frequency between $m - \cdot 075$ and $m + \cdot 075$ half-inches. From these data we deduced m_x and σ_x being in half-inch units.

TABLE X.

Quantity.	1.	2.	3.
Mean, m_x	2·7216	2·6379	2·6445
S.D., σ_x	·4909	·5253	·5072
$m_x/\sigma_x = v_x$	·1804	·1991	·1918

Since $$x'_q = x_q - \tfrac{1}{2}u,$$

we have $$\delta x'_q = \delta x_q - \tfrac{1}{2}\delta u,$$

and $$\sigma^2_{x'_q} = \sigma^2_x + \tfrac{1}{4}\sigma^2_u - \sigma_u\sigma_x r_{ux_q} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(xv.)}.$$

Further, $$\delta x'_q \delta u = \delta x_q \delta u - \tfrac{1}{2}\delta u^2$$

and $$\sigma_{x'_q}\sigma_u r_{ux'_q} = \sigma_{x_q}\sigma_u r_{ux_q} - \tfrac{1}{2}\sigma_u^2,$$

whence $$r_{ux'_q} = \frac{\sigma_{x_q} r_{ux_q} - \frac{1}{2}\sigma_u}{\sigma_{x'_q}} \quad \text{(xvi.)}.$$

Thus (xiv.) gives us r_{ux_q}, (xv.) $\sigma_{x'_q}$, and (xvi.) $r_{ux'_q}$. The numerical values obtained were, σ_x', being in half-inch units :—

TABLE XI.

Quantity.	1.	2.	3.
r_{ux}	·9640	·9514	·9613
$\sigma_{x'}$	·1306	·1635	·1402
$r_{ux'}$	$-$ ·0186 ± ·0302	+ ·1465 ± ·0295	+ ·0851 ± ·0299

Now $r_{ux'_q}$ is the correlation between the absolute error made by the qth observer and the length of the line bisected, and we see at once that, contrary to our *à priori* assumption, there is little relationship between the amount of error and the length of the line bisected. Dr. LEE even makes a larger absolute error for small than for large lines, but her correlation is below its probable error in value, and we can only conclude that the length of the line between the limits taken for it in the experiments is quite immaterial to her judgment of its midpoint. There is a small correlation between Mr. YULE's error and the length of the line, his error increasing if the line be longer. I am the only one of the three experimenters whose judgment of the midpoint of a line is considerably influenced by its length, but even in my case the result is of a totally different order from what we *à priori* had anticipated, for we had supposed the error would be almost directly proportional to the length of the line dealt with.

Clearly, in correlating the judgments of (1) and (2) or of (1) and (3) we should have done better to take absolute displacements of the midpoint, rather than the proportions these bear to the length of the line. Accordingly I proceeded to deduce formulæ for finding the correlations between the absolute displacement errors.

Since $$x'_q = x_q - \tfrac{1}{2}u, \qquad \delta x'_q = \delta x_q - \tfrac{1}{2}\delta u,$$
$$x'_p = x_p - \tfrac{1}{2}u, \qquad \delta x'_p = \delta x_p - \tfrac{1}{2}\delta u,$$

we have, by multiplying out and summing,

$$\sigma_{x'_q}\sigma_{x'_p} r_{x'_q x'_p} = \sigma_{x_q}\sigma_{x_p} r_{x_q x_p} - \tfrac{1}{2}\sigma_u \sigma_{x_q} r_{ux_q} - \tfrac{1}{2}\sigma_u \sigma_{x_p} r_{ux_p} + \tfrac{1}{4}\sigma_u^2.$$

Or $$r_{x'_q x'_p} = \frac{\sigma_{x_q}\sigma_{x_p} r_{x_p x_q} - \frac{1}{2}\sigma_u \sigma_{x_q} r_{ux_q} - \frac{1}{2}\sigma_u \sigma_{x_p} r_{ux_p} + \frac{1}{4}\sigma_u^2}{\sigma_{x'_q}\,\sigma_{x'_p}} \quad \text{(xvii.)}.$$

Now, σ_x, $\sigma_{x'}$, σ_u and r_{ux} are all known. Hence the correlation of the absolute displacements will be known as soon as we find $r_{x_p x_q}$.

But
$$X_q/X_p = x_q/x_p,$$
or,
$$\delta X_q/m_{X_q} - \delta X_p/m_{X_p} = \delta x_q/m_{x_q} - \delta x_p/m_{x_p}.$$

Hence squaring and summing we find:

$$v^2_{X_q} + v^2_{X_p} - 2v_{X_p}v_{X_q}r_{X_pX_q} = v^2_{x_q} + v^2_{x_p} - 2v_{x_p}v_{x_q}r_{x_px_q}.$$

But since $X_p = X'_p + \cdot 5$, $\delta X_p = \delta X'_p$, whence we have at once :—

$$\sigma_{X_p} = \sigma_{X'_p},$$
and
$$r_{X_pX_q} = r_{X'_pX'_q}.$$

Thus we find :—

$$r_{x_px_q} = \tfrac{1}{2}\left\{\frac{v^2_{x_q} + v^2_{x_p}}{v_{x_q}v_{x_p}} - \frac{\dfrac{\sigma^2_{X'_q}}{(m_{X'_q} + \cdot 5)^2} + \dfrac{\sigma^2_{X'_p}}{(m_{X'_p} + \cdot 5)^2} - \dfrac{2\sigma_{X'_q}\sigma_{X'_p}r_{X'_qX'_p}}{(m_{X'_q} + \cdot 5)(m_{X'_p} + \cdot 5)}}{v_{x_q}v_{x_p}}\right\}. \quad . \quad \text{(xviii.)}.$$

Here v_x is given by Table X., $m_{X'}$, $\sigma_{X'}$ and $r_{X'_qX'_p}$ are all entered in Table I., so that $r_{x_px_q}$ can be found. Hence from (xvii.) we find $r_{x_p'x'_q}$ the correlation of the absolute displacements.

Substituting the numerical values we easily find the following results :—

TABLE XII.

$r_{x_2x_3} = \cdot 9445$	$r = \cdot 3596 \pm \cdot 0263$
$r_{x_3x_1} = \cdot 9359$	$r_{x_3x_1'} = \cdot 1242 \pm \cdot 0297$
$r_{x_1x_2} = \cdot 9358$	$r_{x_1'x_2'} = \cdot 2223 \pm \cdot 0287$

There are thus seen to be substantial correlations between the errors in the absolute displacements, not reduced to the length of the bisected line as unit.

To find the correlations between the relative displacements not reduced to the length of the bisected line as unit, we have to find first

$$\sigma^2_{x'_qx'_p} = \sigma^2_{x'_q} + \sigma^2_{x'_p} - 2\sigma_{x'_q}\sigma_{x'_p}r_{x'_qx'_p},$$

using Tables XI. and XII., and thence find

$$\rho_{1,23} = \frac{\sigma^2_{x'_1x'_2} + \sigma^2_{x'_1x'_3} - \sigma^2_{x'_2x'_3}}{2\sigma_{x'_1x'_2}\sigma_{x'_1x'_3}}.$$

There results, the standard deviations being in half-inch units :

TABLE XIII.

$\sigma_{x'_2 x'_3}$ = ·1729	$\rho_{1,23}$ = ·5503 ± ·0210
$\sigma_{x'_3 x'_1}$ = ·1793	$_{2,31}$ = ·5002 ± ·0226
$\sigma_{x'_1 x'_2}$ = ·1852	$\rho_{3,12}$ = ·4478 ± ·0241

We have now the complete data requisite for analysing the experiments on the bisection of straight lines. We place all the correlation coefficients together in Table XIV. for comparison of the two methods of deducing results.

TABLE XIV.—Correlation in the Judgments as to Midpoint of Lines.

	Errors measured in terms of length of line.	Errors measured absolutely.
r_{23}	·3627 ± ·0262	·3596 ± ·0263
r_{31}	·1139 ± ·0298	·1242 ± ·0297
r_{12}	·2053 ± ·0289	·2223 ± ·0287
$\rho_{1,23}$	·5615 ± ·0207	·5503 ± ·0210
$\rho_{2,31}$	·4980 ± ·0227	·5002 ± ·0226
$\rho_{3,12}$	·4379 ± ·0244	·4478 ± ·0241

We conclude at once that :—

(i.) Within the limits of the probable errors of the observations the correlations of the errors in judgment, whether measured absolutely or in terms of the length of the line bisected, are sensibly the same; and this is true not only for the correlation in absolute (r) but also for the correlation (ρ) in relative judgments.

(ii.) Thus while we have shown that the error in bisecting a line is not proportional to the length of the line, and indeed not at all or only slightly correlated with it, yet the observed correlation of judgments cannot arise solely from the use of a ratio or index. For this correlation still exists, if we deal with the absolute errors. It is thus not a purely spurious correlation.

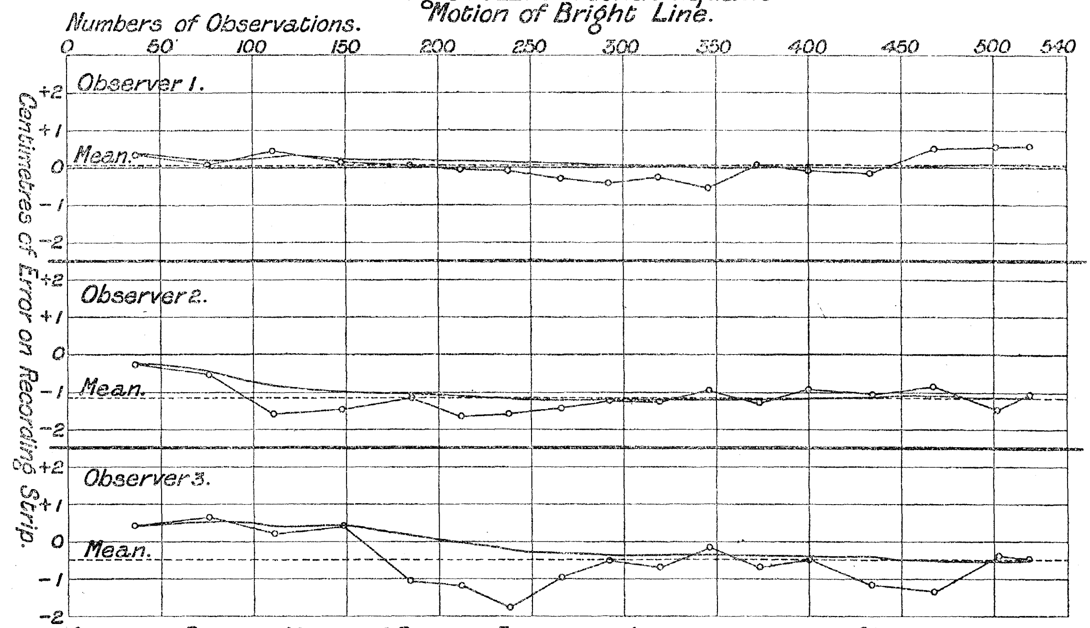
(iii.) The correlation varies considerably from one pair of observers to a second. We thus are forced to conclude that it is not a result of a common varying external cause, but must in part or wholly be due to a common element in the personalities of two experimenters, which is affected in the same way, and differently from some other common element in the personalities of another pair of observers.

Thus the bisection experiments entirely confirm the conclusions we have formed as a result of the bright-line experiments. In both cases there is a real personal correlation of judgment, only in the two series it is differently masked by or combined with spurious correlation according to the special manipulation used in the reduction of the errors.

Taking into account what we have learnt as to the nature of fluctuations in personal equation, I think we may conclude broadly as follows :—

The errors of judgment of apparently independent observers are not as a rule independent. The immediate atmosphere of each single observation or of each short series of observations affects in a differential manner the factors of the personality, causing variations in the personal equation which are not of the order of those due to random sampling. Certain factors affected by the immediate atmosphere seem to be common elements of two or more personalities, and there results from this a tendency in each pair of observers to judge in the same manner. If we enlarge the concept of "immediate atmosphere" to embrace not only the objective side of the phenomena, but the physical and mental state of the percipient, we may simply state that certain elements of this immediate atmosphere are common to each pair of observers and produce a correlation between their judgments. Their personal equations fluctuate in sympathy. This sympathetic fluctuation of personal equations leading to correlation of judgments is really visible on inspection, as the reader will at once see on examining Diagrams II. and III.

Diagram III. Personal Equation Motion of Bright Line.



Continuous Polygon = Values of Personal Equation as based upon 27 to 37 Observations.
Continuous Curve = Approach " " " to its Mean Value for 519 " .

This quite sub-conscious sympathetic fluctuation of personal equation in the case of apparently independent observers is not only of fundamental importance when we have to combine observations of the same phenomenon by different observers, and assign the weight of the combination, but it appears to have an even wider bearing when we have to consider to what degree the testimony of a number of apparently independent witnesses of the same event is strengthened by the concurrence of their judgments as to what actually took place. Without some estimate of the correlation of judgments we cannot assert what weight is to be given to combined testimony.

(9.) *On the Nature of the Frequency Distribution in the case of Errors of Judgment.*

Having completed our investigation of the nature of fluctuations in personal equation and of the correlation between judgments—an investigation which demands no hypothesis as to the form of their law of distribution—we now turn to a consideration of the manner in which errors of judgment are distributed.

In Tables XV. and XVI. will be found the frequencies for the two series of experiments, the results being *grouped* (see pp. 272–273).*

The question to be answered is this: Is the general nature of these distributions capable of being described by the "normal" curve of errors, on the assumption that they are random samplings of the whole "populations" of errors that the observers respectively would produce if they continued to experiment indefinitely under the same conditions? So far as I am aware no thorough investigation has yet been made as to how far actually observed errors are capable of being described by the normal curve of errors. In most text-books on the theory of errors certain axioms are laid down as ruling the distribution of errors of judgment, and on the basis of these axioms the normal curve of errors is deduced. One or two limited series of errors of observation are then cited, and the axioms declared to be satisfactory by comparing a graph of the theoretical with the observed distribution, or by a table comparing the observed and theoretical frequencies of errors occurring within each small range. As a rule a vague inspection of the amount of agreement is the only thing appealed to to test the accordance of theory and experiment. So far as I am aware writers on the theory of errors have quite overlooked the point that that theory itself provides a perfectly general test of whether the accordance between theory and experiment is a reasonable or an unreasonable one. It is not a question of whether there is a "practical accordance" between the two, whatever that may mean, but of the degree of probability that a given system of errors or deviations is a random sampling from an indefinitely large distribution of errors obeying the axioms from which the normal curve of errors has been deduced. To talk of "practical accordance" between theory and observation is simply to shuffle out of an examination of the truth, when the odds are 3000 to 1, or even 70 to 1, against the observed results being a random sample of errors obeying certain fundamental axioms.† Now in the

* In Table XV. a group such as 4·755 embraces all the frequency between 4·505 and 5·005; and in Table XVI. a group such as ·04 embraces all the frequency between ·035 and ·045.

† A recent writer on statistics seems to find that an agreement measured by the odds of 3000 to 1 is *very satisfactory*, and one against which the odds are 70 to 1 represents with all *practicable accuracy* the observ ed frequency. Comment is needless.

TABLE XV.—Frequency of Absolute and Relative Errors in Bright-line Series of Experiments. Number, 519

Size of error.	(1.)	(2.)	(3.)	(3–2.)	(1–3.)	(2–1.)
7·755	—	—	—	1	—	—
7·255	—	—	—	1	—	—
6·755	—	—	—	—	1	—
6·255	1	—	1	—	1	—
5·755	—	—	—	4	2	—
5·255	—	1	—	1	5	—
4·755	1	—	—	4	5	—
4·255	—	—	—	12	17	—
3·755	1	—	1	15	18	—
3·255	6	—	12	19	17	1
2·755	4	—	19	20	29	1
2·255	12	—	18	26	35	8
1·755	22	3	26	34	41	13
1·255	57	8	42	49	37	19
·755	71	31	48	62	41	23
·255	97	35	44	77	60	54
− ·245	85	73	48	90	41	72
− ·745	69	76	46	41	56	55
− 1·245	56	96	60	21	35	50
− 1·745	23	79	42	19	34	59
− 2·245	7	60	36	9	21	58
− 2·745	4	30	36	5	10	38
− 3·245	1	17	20	5	6	30
− 3·745	1	5	12	1	5	15
− 4·245	—	3	5	3	2	7
− 4·745	1	1	2	—	—	7
− 5·245	—	—	1	—	—	3
− 5·745	—	—	—	—	—	2
− 6·245	—	—	—	—	—	2
− 6·745	—	1	—	—	—	—
− 7·245	—	—	—	—	—	—
− 7·745	—	—	—	—	—	1
− 8·245	—	—	—	—	—	—
− 8·745	—	—	—	—	—	—
− 9·245	—	—	—	—	—	—
− 9·745	—	—	—	—	—	—
− 10·245	—	—	—	—	—	1

present investigation we have no less than twelve frequency distributions, six absolute distributions and six relative distributions; the latter being of course of the type which will usually occur in astronomical or physical observations where the absolute errors cannot be measured. We have then material *enough* to discuss the problem: Is it suitable for the purpose? It seems to me that there is nothing peculiar to our data which marks them off from other series of observational errors, except their rather extensive character, which was necessary if safe conclusions were to be drawn. There were four independent observers, three of whom at least had been long used to making observations and measurements; the fourth, less accustomed, turned out in the sequel to have the steadiest judgment. Further, the investigations were begun with

TABLE XVI.—Frequency of Absolute and Relative Errors referred to Length of Line as Unit in Bisection Experiments.* Number = 500.

Size of error.	(1.)	(2.)	(3.)	(2–3.)	(3–1.)	(1–2.)
−·12	—	—	—	—	—	2
−·11	—	—	—	—	—	1
−·10	—	—	—	1	—	3
−·09	1	—	—	2	—	7
−·08	4	1	—	1	1	9
−·07	8·5	3	1	6	2	23·5
−·06	12	11	7·5	13	6·5	21·5
−·05	13·5	14·5	9·5	16·5	14·5	35
−·04	45	21·5	22	26	10	58
−·03	61	30	40·5	45·5	26	45
−·02	76	47	43·5	36	43	59
−·01	90·5	51·5	51	66	33	64·5
·00	74·5	72	68·5	55·5	42	43
+·01	50	65·5	75	61	53·5	37·5
+·02	30·5	53	70·5	52·5	60·5	26·5
+·03	21·5	50·5	61	41·5	61	27·5
+·04	7	28·5	25·5	34·5	52	16
+·05	3	27	13·5	20	34·5	7
+·06	2	13·5	10	13	27	11
+·07	—	7·5	1	4	17·5	1
+·08	—	—	—	2	5	1
+·09	—	1	—	2	7·5	1
+·10	—	—	—	1	3·5	—
+·11	—	2	—	—	—	—

no intention of considering the problem of normal frequency; they were designed to demonstrate what appeared a remarkable and valuable result flowing from the theory of errors as usually expounded (see p. 240). Each of us made our individual judgments with care and without any theoretical bias. We were of course, during the work of the observations, liable to physical and psychological influences, to the subtle changes of daily health and of sense-keenness. But I contend that all such things affect every observer, and that it is idle to propound a theory which would hold for an ideal observer of perfectly equable temperament and physical fitness observing under a perfectly equable environment for a number of days or even weeks an exactly identical phenomenon. Such a theory could not be verified, and if verified would have no practical application. Our observations seem to me a perfectly fair sample of actual errors of judgment, and I believe no objections can be taken to them which would not apply with even increased force against most of the series of errors of judgment with which physicists or astronomers have to deal.

There are two classes of considerations which arise in our view of frequency distributions :—

* In this table and in the diagrams X. to XV. the error has been given the opposite sign to its value in Table I.

(*a*.) General physical characters of the nature of the distribution without regard to the special frequency of errors of particular sizes.

(*b*.) Agreement between theory and observation in the general distribution of errors of each particular size.

I propose to investigate these classes of considerations separately.

(10.) (*a*.) *General Physical Characters of a Normal Distribution.*

While the analytical processes by means of which the normal curve is deduced are extremely varied—sometimes very simple (HAGEN), sometimes very complex (POISSON), there is confessedly or tacitly involved an axiom of the following kind :—

(α.) *Positive and negative errors of the same size are equally frequent.* Sometimes this result is disguised by assuming that the actual error is the sum of an indefinitely great number of small elementary errors which are equally likely to be positive or negative. Whatever process of proof be followed the result is the same—the normal distribution gives a symmetrical distribution of errors, and this is its first general physical character. Now in an immense number of cases of deviations from the mean, such as occur in organic nature, this symmetry is quite unknown ; such distributions I have spoken of as skew frequency distributions,* and their characteristic feature is that the mode or position of the maximum frequency diverges from the mean. The ratio of the distance of the mode from the mean to the standard deviation I have treated as a measure of the "skewness" of the distribution. It will vanish when the curve is symmetrical or when the sums of all odd powers of the errors are zero. Thus if n be the number of observations, $n\mu_p$ the sum of the pth powers of the errors, $\mu_p = 0$ for a normal distribution if p be odd. For most practical purposes the labour of investigation compels us to confine our attention to the question of whether μ_3 is sensibly zero.

But the normal distribution not only involves a condition as to the odd moments, but also one which must sensibly hold in the case of each pair of even moments.† The simplest of such relations is expressed by

$$\mu_4 = 3\mu_2^2.$$

Now this relation and its extensions to higher moments have nothing whatever to do with the symmetry of the normal distribution—with the equal frequency of errors of the same size, whether positive or negative. They depend really upon two additional axioms, which are again confessedly or tacitly assumed in the course of the proof, namely :—

(β.) That there are an indefinitely great number of cause-groups associated in producing each individual error.

* 'Phil. Trans.,' A, vol. 186, p. 343 *et seq.*

† 'Phil. Trans.,' A, vol. 185, p. 108.

(γ.) That the contributions towards any individual error of these cause-groups are not correlated among themselves.

It is not my purpose at present to consider the philosophical arguments for or against these axioms. I have considered the matter at length in a paper not yet published, but I want to indicate the source of such relations as those just referred to. If actual distributions of error do not sensibly satisfy $\mu_3 = 0$, then axiom (α) is not true; if they do not sensibly satisfy $\mu_4 = 3\mu_2^2$, then either (β) or (γ) or both are invalid.

The nature of the relation $\mu_4 = 3\mu_2^2$ deserves a little fuller consideration from the physical side.

Let m_1 and m_2 be the number of errors of magnitudes x_1 and x_2 respectively, and let $n'\mu_2'$ and $n'\mu_4'$ represent the second and fourth moments of the remainder of the errors. Let $m_1 + m_2 = m'$. Then

$$n\mu_2 = n'\mu_2' + m_1x_1^2 + m_2x_2^2,$$
$$n\mu_4 = n'\mu_4' + m_1x_1^4 + m_2x_2^4.$$

Hence we find :—

$$n\mu_4 = n'\mu_4' + m'x_2^4 + (n\mu_2 - n'\mu_2' - m'x_2^2)(x_1^2 + x_2^2).$$

Now without altering the total frequency, *i.e.*, keeping m' constant, take part of the frequency m_2 and transfer it from x_2 to x_1; do this equally on both sides of the mean, so that the position of the mean be not altered. Now in order that μ_2 should also not be altered, x_2 being supposed constant, we must have :—

$$\delta m_1/m_1 = (2x_1\delta x_1)/(x_2^2 - x_1^2),$$

or if x_2 be $> x_1$, δx_1 must be positive. Thus if we bring a part of the outlying frequency inward to a point nearer the mean, we can still retain the same mean and the same standard deviation, *i.e.*, get the same normal curve, if we shift the inlying frequency group a little outward. The whole effect of such a change will be to flatten the frequency curve at its summit by a reduction of its tails, which increases the middle part of the curve. Now, looking at the above value of μ_4 we see that since $x_1^2 < x_2^2$, $n\mu_2 - n'\mu_2' - m'x_2^2$ is negative, and therefore that when x_1 increases μ_4 decreases.

We conclude accordingly that symmetrical or nearly symmetrical curves which have the same mean and standard deviation as a normal curve will be flatter topped if μ_4 be $< 3\mu_2^2$, and steeper at the top if μ_4 be $> 3\mu_2^2$.

Take, for example, the details of shots at a target given by MERRIMAN, 'Method of Least Squares,' p. 14: here*

$$\mu_2 = 2{\cdot}402{,}343, \qquad \mu_4 = 14{\cdot}578{,}491,$$

and $$3\mu_2^2 = 17{\cdot}313{,}752.$$

* Using SHEPPARD'S corrective terms, 'London Math. Soc. Proc.,' vol. 29, p. 369.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Accordingly μ_4 is $< 3\mu_2^2$ by a considerable amount, and the observations are immensely flatter than the normal curve with which they can be fitted. Actually the normal curve has a maximum ordinate which rises some 15 to 20 per cent. above the corresponding ordinate of the observations. Hence quite apart from the question of equal negative and positive errors, we should assert that because $\mu_4 = 3\mu_2^2$ is not sensibly satisfied, it follows that one or other or both the axioms (β), (γ) cannot be true for this distribution of hits.

I propose to look a little more closely into the probable errors of the quantities connected with a normal distribution. I take d to be the distance from the mean to the mode, and define the skewness, Sk., as in earlier memoirs, to be the ratio of d to σ. I write $\beta_1 = \mu_3^2/\mu_2^3$ and $\beta_2 = \mu_4/\mu_2^2$. Mr. FILON and I have dealt with the probable errors of skew frequency curves in a special memoir,* and deduced those for the normal curve as the limit to those of a skew curve of what I have termed Type III. The disadvantage of this procedure is that it supposes the deviations from symmetry to take place along a class of curve for which

$$2\mu_2(3\mu_2^2 - \mu_4) + 3\mu_3^2 = 0 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(xix.)}.$$

μ_4 is thus known in terms of μ_2 and μ_3. The result is that when μ_3 is put zero to reach the normal case, the error of μ_4 is found to be absolutely correlated with that of μ_2, and the probable value of this error to be deducible from that of μ_2 by means of the relation

$$\mu_4 = 3\mu_2^2.$$

To obtain perfectly general results we must use not Type III., but Type I. or Type IV. of that memoir, curves in which μ_2, μ_3, and μ_4 are absolutely independent of each other. Our results can easily be deduced by aid of certain elegant formulæ due to Mr. W. F. SHEPPARD.† In our notation these are:—

Probable error of $\mu_p = \cdot 67449\, \Sigma_{\mu_p}$ is given by

$$\Sigma^2_{\mu_p} = \frac{\mu_{2p} - \mu^2_p + p^2 \mu^2_{p-1}\mu_2 - 2p\mu_{p+1}\mu_{p-1}}{n} \quad . \quad . \quad . \quad . \quad . \quad \text{(xx.)}.$$

$R_{\mu_p \mu_q}$ = Correlation of errors in μ_p and μ_q is given by

$$\Sigma_{\mu_p}\Sigma_{\mu_q} R_{\mu_p\mu_q} = \frac{\mu_{p+q} - p\mu_{p-1}\mu_{q+1} - q\mu_{p+1}\mu_{q-1} + pq\mu_{p-1}\mu_{q-1}\mu_2 - \mu_p\mu_q}{n} \quad . \quad \text{(xxi.)}.$$

These results are perfectly general whatever be the law of the frequency. As special cases we have, when $\mu_1 = 0$:

* 'Phil. Trans.,' A, vol. 191, pp. 229–311, especially p. 276.

† 'Phil. Trans.,' A, vol. 192, p. 126.

$$\left.\begin{aligned}
\Sigma^2_{\mu_2} &= (\mu_4 - \mu_2^2)/n \\
\Sigma^2_{\mu_3} &= (\mu_6 - \mu_3^2 + 9\mu_2^2 - 6\mu_4\mu_2)/n \\
\Sigma^2_{\mu_4} &= (\mu_8 - \mu_4^2 + 16\mu_3^2\mu_2 - 8\mu_5\mu_3)/n \\
\Sigma_{\mu_2}\Sigma_{\mu_3}\mathrm{R}_{\mu_2\mu_3} &= (\mu_5 - 4\mu_3\mu_2)/n \\
\Sigma_{\mu_3}\Sigma_{\mu_4}\mathrm{R}_{\mu_3\mu_4} &= (\mu_7 - 3\mu_2\mu_5 - 5\mu_4\mu_3 + 12\mu_2^2\mu_3)/n \\
\Sigma_{\mu_2}\Sigma_{\mu_4}\mathrm{R}_{\mu_2\mu_4} &= (\mu_6 - 4\mu_3^2 - \mu_2\mu_4)/n
\end{aligned}\right\} \quad . \quad . \quad . \quad \text{(xxii.)}.$$

For the special case of the normal curve, since

$$\mu_4 = 3\mu_2^2 = 3\sigma^4, \quad \mu_6 = 15\sigma^6, \quad \mu_8 = 105\sigma^8$$
$$\mu_7 = \mu_5 = \mu_3 = 0,$$

we have :—

$$\text{Probable error of } \mu_2 = \cdot 67449 \times \sqrt{2}\sigma^2/\sqrt{n} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(xxiii.)}.$$

$$\text{,, \quad ,, \quad } \mu_3 = \cdot 67449 \times \sqrt{6}\sigma^3/\sqrt{n} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(xxiv.)}.$$

$$\text{,, \quad ,, \quad } \mu_4 = \cdot 67449 \times \sqrt{96}\sigma^4/\sqrt{n} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(xxv.)}.$$

$$\mathrm{R}_{\mu_2\mu_3} = 0 \quad . \quad . \quad \text{(xxvi.)}. \qquad \mathrm{R}_{\mu_3\mu_4} = 0 \quad . \quad . \quad \text{(xxvii.)}. \qquad \mathrm{R}_{\mu_2\mu_4} = \tfrac{1}{2}\sqrt{3} \quad . \quad . \quad \text{(xxviii.)}.$$

In a further memoir on skew variation,* not yet published, I show that if the differential equation to the frequency curve be

$$\frac{1}{y}\frac{dy}{dx} = \frac{a_0 + a_1 x}{b_0 + b_1 x + b_2 x^2} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(xxix.)},$$

then whatever be the form of the curve, the distance d between the mean and the mode, and the skewness are always given by

$$d = \frac{\mu_3(\mu_4 + 3\mu_2^2)}{2(5\mu_2\mu_4 - 6\mu_3^2 - 9\mu_2^3)} = \frac{1}{2}\frac{\sqrt{\mu_2}\sqrt{\beta_1}(\beta_2 + 3)}{5\beta_2 - 6\beta_1 - 9} \quad . \quad . \quad . \quad . \quad \text{(xxx.)}.$$

$$\text{Sk.} = \frac{\mu_3(\mu_4 + 3\mu_2^2)}{2\sqrt{\mu_2}(5\mu_2\mu_4 - 6\mu_3^2 - 9\mu_2^3)} = \frac{1}{2}\frac{\sqrt{\beta_1}(\beta_2 + 3)}{5\beta_2 - 6\beta_1 - 9} \quad . \quad . \quad . \quad . \quad \text{(xxxi.)}.$$

* My original memoir ('Phil. Trans.,' A, vol. 186, p. 343), being much misunderstood, has been alternately over- and under-rated. I had found that the ordinary theory of errors was far from describing frequencies within the limits of error imposed by a random sampling. My object was then to discover a series of curves which would enable me in a very great number of cases to do this. I did not select these curves at random, but endeavoured to see where the usual hypotheses failed and must be generalised. My hypergeometrical series was not empirically chosen, but on the grounds of the axioms (α), (β), (γ) above. I chose a system where the positive and negative errors were not equally probable, where there was not an infinite number of cause-groups, and lastly, one where these cause-groups did not contribute independent but correlated elements to the total error. All these points as well as criticisms, mostly due to complete misunderstanding of what random sampling means, I have considered in a further memoir.

Hence, since β_1 and β_2 can always be found, we have general expressions for d and Sk. in the case of the first four moments being arbitrary.

Further, if we increase our β series and write :—

$$\beta_1 = \mu_3^2/\mu_2^3, \quad \beta_2 = \mu_4/\mu_2^2, \quad \beta_3 = \mu_3\mu_5/\mu_2^4,$$
$$\beta_4 = \mu_6/\mu_2^3, \quad \beta_5 = \mu_7\mu_3/\mu_2^5, \quad \beta_6 = \mu_8/\mu_2^4 \quad \text{(xxxii.)},$$

we have the following perfectly general results, the law of frequency being any whatever :—

$$n\Sigma_{\beta_1}^2 = \beta_1(4\beta_4 - 24\beta_2 + 36 + 9\beta_1\beta_2 - 12\beta_3 + 35\beta_1) \quad \text{(xxxiii.)}.$$

$$n\Sigma_{\beta_2}^2 = \beta_6 - 4\beta_2\beta_4 + 4\beta_2^3 - \beta_2^2 + 16\beta_2\beta_1 - 8\beta_3 + 16\beta_1 \quad \text{(xxxiv.)}.$$

$$\Sigma_{\beta_1}\Sigma_{\beta_2}R_{\beta_1\beta_2} = 2\beta_5 - 3\beta_4\beta_1 - 4\beta_3\beta_2 + 6\beta_2^2\beta_1 + 3\beta_1\beta_2 - 6\beta_3 + 12\beta_1^2 + 24\beta_1 \quad \text{(xxxv.)}.$$

Lastly if

$$K = 6 + 3\beta_1 - 2\beta_2 \quad \text{(xxxvi.)}.$$

$$n\Sigma_K^2 = 4\beta_6 - 16\beta_2\beta_4 + 16\beta_2^3 + 72\beta_1\beta_4 - 24\beta_5 - 72\beta_1\beta_2^2 + 48\beta_2\beta_3 + 81\beta_1^2\beta_2$$
$$- 108\beta_1\beta_3 - 4\beta_2^2 - 188\beta_1\beta_2 + 72\beta_3 + 171\beta_1^2 + 100\beta_1 \quad \text{(xxxvii.)}.$$

Thus the probable errors $\cdot 67449\ \Sigma_{\beta_1}$, $\cdot 67449\ \Sigma_{\beta_2}$, $\cdot 67449\ \Sigma_K$ of the quantities β_1, β_2, and what I have termed the criterion K, can be found *whatever be the law of frequency.*

Let $\epsilon_1 = \sqrt{\beta_1} = \mu_3/(\mu_2)^{\frac{3}{2}}$, then

$$\Sigma_{\epsilon_1} = \tfrac{1}{2}\Sigma_{\beta_1}/\sqrt{\beta_1} \quad \text{(xxxviii.)},$$

and its value can be found from (xxxiii.). Knowing Σ_{β_1}, Σ_{β_2}, and the correlation of errors in β_1 and β_2, *i.e.*, $R_{\beta_1\beta_2}$, we can find at once the probable errors in d and Sk. from (xxx.) and (xxxi.). Thus with very great generality as to the law of frequency, we can test how far the distribution is a random sample from a population following any law whatever.

Applying the above general results to the special case of the normal curve, we find since

$$\beta_1 = 0, \quad \beta_2 = 3, \quad \beta_3 = 0, \quad \beta_4 = 15, \quad \beta_5 = 0, \quad \beta_6 = 105,$$

Probable error of $\beta_1 = 0$

„ „ $\beta_2 = \cdot 67449 \times \sqrt{\dfrac{24}{n}}$ . . . . (xxxix.),

„ „ $\sqrt{\beta_1} = \cdot 67449 \times \sqrt{\dfrac{6}{n}}$ . . . . (xl.),

$R_{\beta_1\beta_2} = 0$

„ „ $K = \cdot 67449 \sqrt{\dfrac{96}{n}}$ . . . . (xli.),

„ „ $d = \cdot 67449 \sqrt{\dfrac{3}{2n}}\,\sigma$ . . . . (xlii.),

„ „ Sk. $= \cdot 67449 \sqrt{\dfrac{3}{2n}}$ . . . . (xliii.).

The result that the probable error of β_1 is zero, but that of $\sqrt{\beta_1}$ is finite, may appear paradoxical, but it is due to the fact that the errors are treated as small quantities and β_1 involves μ_3^2, the square of a small quantity, or one zero, if the distribution were truly normal.

Of these results, Mr. FILON and I have already published with a less general proof (xxiii.), (xxiv.), (xxvi.), (xxvii.), (xlii.), and (xliii.). Instead of (xxv.) and (xxviii.), we found

$$\text{Probable error of } \mu_4 = \cdot 67449 \sqrt{72}\sigma^4/\sqrt{n}, \text{ and } R_{\mu_2\mu_4} = 1,$$

i.e., replacing $\sqrt{96} = 9\cdot 8$ by $\sqrt{72} = 8\cdot 5$, and $\cdot 866$ by 1. This was due to the fact that we considered the variations from normality to be given by a distribution of Type III. (see p. 276). The differences, however, are not such as to invalidate arguments based on the general order of the probable error of μ_4.*

We have now general relations enough to answer the following questions :—

(i.) Does the value of d found from (xxx.) differ from zero by an amount large as compared with the probable error of d given in (xlii.)?

(ii.) Does the skewness found from (xxxi.) differ from zero by an amount large as compared with the probable error of the skewness as given in (xliii.)?

(iii.) Does the value of μ_3 differ from zero by a value large as compared with the probable error of μ_3 given in (xxiv.)?

In all these questions we have a test of whether the distribution is really a random selection from a symmetrical distribution, *i.e.*, from material obeying axiom (α). The same thing is again dealt with by testing the error of $\sqrt{\beta_1}$ as given by (xl.).

(iv.) Is the condition $\mu_4 = 3\mu_2^2$, or $\beta_2 = 3$ satisfied for the distribution, *i.e.*, does β_2 differ from 3 by a quantity which is not large as compared with its probable error as given by (xxxix.)?

If (i.) to (iii.) are satisfied, but not (iv.), the distribution is still not a random selection from material obeying the normal law, *i.e.*, axioms (β) and (γ) cannot both be true for it.

Lastly, if the material does not obey the normal law, does the criterion K differ sensibly from zero, and therefore form a characteristic to be regarded?

I turn first to the motion of the bright line and give in Table XVII. the constants for this series of distributions.

* See 'Phil. Trans.,' A, vol. 191, p. 277.

TABLE XVII.—Motion of Bright Line.

	1.	2.	3.	3–2.	1–3.	2–1.
μ_2	5·6561	5·5027	13·2709	12·0942	16·1136	11·1019
μ_3	+4·7424 ± ·9755	+ ·5452 ± ·9361	+5·5761 ±3·5060	+21·0453 ±3·0502	+16·2877 ±4·6909	−20·4072 ±2·6827
μ_4	160·3883	152·1762	456·8883	583·8991	693·2790	569·4060
β_1	·1243 ·0000	·0018 ·0000	·0133 ·0000	·2504 ·0000	·0634 ·0000	·3044 ·0000
β_2	5·0135 ± ·1450	5·0256 ± ·1450	2·5942 ± ·1450	3·9919 ± ·1450	2·6701 ± ·1450	4·6918 ± ·1450
d	·2193 ± ·0862	·0247 ± ·0851	·3020 ± ·1321	·6432 ± ·1261	·7219 ± ·1456	·5706 ± ·1088
Sk.	·0922 ± ·0363	·0105 ± ·0363	·0829 ± ·0363	·1850 ± ·0363	·1798 ± ·0363	·1713 ± ·0363
$\sqrt{\beta_1}$	·3524 ± ·0725	·0422 ± ·0725	·1153 ± ·0725	·5004 ± ·0725	·2518 ± ·0725	·5517 ± ·0725
K	−3·6541 ± ·2901	−4·0460 ± ·2901	·8514 ± ·2901	−1·2327 ± ·2901	·8501 ± ·2901	−2·3266 ± ·2901

N.B.—The units of this table are $\frac{1}{2}$ centim. on the recording strip—not on the observation strip; these are the units of our grouping in Table XV. In most of my previous tables the unit has been taken as 1 centim. of the recording strip. The $\frac{1}{2}$ centim. is retained here, as it will be required in the plotted diagrams as unit of grouping.

Now let us examine these results, remembering that on the basis of a random sampling the odds against a quantity exceeding its supposed value by twice, thrice, four, five times its probable error, are 10 to 1, 49 to 1, 332 to 1, 2700 to 1 respectively, in round numbers.

In the first place, μ_3 differs from zero by 4 to 6 times the probable error in (1), (3–2), (1–3), and (2–1). Further, d differs from zero by 2·5 to 5 times the probable error in the same cases, and the skewness also by 2·5 to 5 times its probable error. $\sqrt{\beta_1}$ differs from zero by 3 to nearly 8 times its probable error in the same four distributions. I consider that it is really impossible to look upon these distributions as random samplings from symmetrical material. On the other hand, (2) and (3) or the absolute personal equation of Dr. MACDONELL and Dr. LEE might well be *symmetrical* distributions. Do they, however, fulfil the conditions for normality $\beta_2 = 3$, $K = 0$? The deviations of β_2 from 3 are in the two cases 2·0256 and 2·4058, or nearly 14 and 2·8 times the probable errors respectively. Further, their values of K differ from zero by nearly 14 and 2·9 times their probable errors. Thus the odds are enormous against Dr. MACDONELL'S judgment being a random sampling from a normal distribution of errors, and are about 300 to 1 against Dr. LEE'S being such! Of the other distributions the odds are enormously against normal distribution in cases (1), (3–2), and (2–1). They are less marked in (1–3), β_2 only differing from 3 and K from 0 by about 2·1 and 2·9, their probable errors respectively—but this case has already been excluded from normality on account of its sensible skewness.

Thus while two of the cases might reasonably be considered symmetrical distributions, *i.e.*, to fulfil axiom (α), and one of the cases might with some improbability (10 or 11 to 1) be supposed to have $\beta_2 = 3$, *i.e.*, to fulfil axioms (β) and (γ), no single case can be supposed, with any reasonable degree of probability, to fulfil all three, or to be capable of representation by a normal curve. The third moment, the distance from mean to mode, the skewness and the magnitude and sign of the criterion K are quantities which, one or more or all, are in each individual case sensible and quite inconsistent with the result of random sampling from "normal material."

I now give a similar table for the bisection experiments.

TABLE XVIII.—Bisection of Lines.

	1.	2.	3.	2–3.	3–1.	1–2.
μ_2	6·0258	9·3966	6·8915	10·4745	11·3987	12·3817
μ_3	−1·7326	·9779	−3·8272	2·2447	−4·6402	2·6583
	±1·0929	±2·1283	±1·3367	±2·5048	±2·8435	±3·2191
μ_4	118·0048	267·1088	124·7219	301·1530	352·0688	450·2880
β_1	·0137	·0012	·0448	·0044	·0145	·0037
	·0000	·0000	·0000	·0000	·0000	·0000
β_2	3·2499	3·0251	2·6261	2·7448	2·7096	2·9372
	± ·1478	± ·1478	± ·1478	± ·1478	± ·1478	± ·1478
d	+ ·1254	− ·0512	+ ·4045	·1310	+ ·2605	− ·1125
	± ·0907	± ·1132	± ·0970	± ·1196	± ·1247	± ·1300
Sk.	·0511	·0006	·1541	·0405	·0772	·0511
	± ·0369	± ·0369	± ·0369	± ·0369	± ·0369	± ·0369
$\sqrt{\beta_1}$	·1171	·0340	·2115	·0662	·1206	·0610
	± ·0739	± ·0739	± ·0739	± ·0739	± ·0739	± ·0739
K	− ·4587	− ·0468	+ ·8821	+ ·5235	+ ·6243	+ ·1368
	± ·2955	± ·2955	± ·2955	± ·2955	± ·2955	± ·2955

Now it will be seen by a most cursory glance at this table that the distribution of errors in the case of the bisection of right lines can be far more nearly represented by a normal curve than in the case of judgment as to the position of a bright line. In the case of every one of the constants for the distribution of my own errors of judgment (*i.e.*, (2)), they differ by less than their probable error from their value on the normal theory. I can therefore treat my judgments as following the normal law and represent them by this curve. In Mr. YULE's case (*i.e.*, (3)), the distance from his mode to his mean is more than four times its probable error; or, only once in 332 trials, say, should we expect such a divergence from normality in a random selecting. It is thus very improbable that his judgments follow the normal law so far as symmetry is concerned. Further, the value of β_2 differs from 3 by about 2·53 times its probable error, or the odds against such a value are about 22 to 1, or, since β_2 can, unlike d and Sk., differ from its normal value either in excess or defect, say 10 or 11 to 1. On both counts, then, Mr. YULE's judgments form improbably a normal distribution. Lastly, turning to Dr. LEE'S (*i.e.*, (1)), we see that the greatest divergence from normality is about 1·7 times the probable error in the case of β_2. The divergence from symmetry is measured by about 1·5 times the probable error. Thus the odds against such a random selection are on the two counts 4 to 1 and 3 to 1, about, or since the two results, given normality, are independent, the combination gives about 12 to 1. These are certainly only moderately long odds, and we must conclude that though a skew curve describes Dr. LEE'S judgment considerably better than a normal curve, yet a normal curve might, without any great improbability, be adopted.

The results for the absolute judgments indicate what we may expect to find for the relative judgments. The relative judgment of Dr. LEE and myself can well be described by a normal curve (see 2–1); the constants differing by less or by very little more from their normal values than the probable errors. On the other hand, Mr. YULE'S and Dr. LEE'S (see 1–3) relative judgments differ from normality on the score both of asymmetry and of flat-toppedness, by odds of more than 10 to 1 in each case, or of 100 (or at least 50) to 1 in the combination. Lastly, the odds against Mr. YULE'S and my relative judgments (see 3–2) on the basis of a random distribution are only about 4 to 1 on the more unfavourable way of considering them, so that 3–2 might pass as a normal distribution.

We thus conclude that while two out of the six distributions in the bisection series are very improbably random selections from normal material, two others are capitally represented by normal curves, while the remaining two are not very favourable cases.

Taking these results in connection with those for the bright-line distributions we must conclude: That the distribution of errors of judgment can diverge in a very sensible way, both on account of asymmetry and of flat-toppedness, from the Gaussian curve of errors; but that cases can be found which approximate with all probability to random sampling from normal material. Consequently it is necessary to select a type or types of frequency curve which, while allowing for these points of sensible divergence, will yet pass into the normal distribution in certain special cases where within the limits prescribed by the probable error the skewness and $\beta_2 - 3$ are sensibly zero.

Since it is incontestible that, if axioms (α), (β), and (γ) are adopted, our distribution of errors must be normal, we must conclude that one or other or all of these axioms are not universally true. When therefore we get material for which the skewness is sensibly not zero, or β_2 is sensibly not three, we are quite at liberty to assert that the sources producing these errors do not fulfil axiom (α) or axioms (β) or (γ) respectively.*

* It is very necessary to insist upon this. A recent critic has asserted that I have argued in the second memoir of my evolution series ('Phil. Trans.,' A, vol. 186, p. 343) in an illegitimate manner on the nature of the sources which lead to particular types of distribution. He denies that it is possible to state anything

(11.) (*b.*) *Agreement between Theory and Observation in the General Distribution of Errors of each particular size.*

Now in the 'Philosophical Magazine' for July, 1900, I have worked out a very simple criterion for the goodness of fit of any frequency distribution to a theoretical curve. I have measured the probability that the divergence from a given curve is one which may be attributed to random sampling. The test is of the following kind: Calculate the squares of the differences of the observed and theoretical frequencies, and divide each such square by the corresponding theoretical frequency; the sum of all such results, written χ^2, is the constant from which we can easily determine whether the probability, P, is large or small that the observed system of divergences or a still more divergent system would arise by random sampling. In Table XIX. below are recorded for the case of the bright-line experiments the values of χ^2, n', or the number of frequency groups, and P, the above-mentioned probability.

TABLE XIX.—Motion of Bright Line.

		1.	2.	3.	3–2.	1–3.	2–1.
Skew curve	n'	18	16	20	24	23	23
	χ^2	12·07	19·72	15·88	60·24	20·37	40·17
	P	·7959	·1829	·6653	·000,035	·5599	·0103
Normal curve	n'	18	16	20	24	24	23
	χ^2	42·85	83·50	21·82	154·41	34·46	99·79
	P	·0006	·000,000	·2933	·000,000	·0441	·000,000

Some words are necessary as to the meaning of this table. n' gives the number of groups of frequency upon which the determination of χ^2 was based. This had to be somewhat arbitrary when there were outlying observations, as in cases (2), (3), (2–1). The calculation of the frequencies within the range of each group was found partly by mechanical integration of carefully drawn diagrams of the

as to these sources. When one advances into a new country one is apt not to see all things at once in their due proportions, and I may well have laid more stress than was justifiable on the importance of range, for example. This was not because a determination of range, if it exists, is not of most primary importance, but because I had not till the fourth memoir of the series ascertained a method of determining the probable error of the determination of range, and seen that in certain cases it is considerable. The critic—to whom I hope to reply elsewhere—seems to have failed to perceive the aim of my investigations *i.e.*, to find a simple description of frequency, which will describe the great bulk of cases *within the errors of random sampling.*

theoretical curves, partly by reading ordinates where the contour of the curves was nearly straight, and partly by quadrature near the tails. Till tables have been calculated for the skew curves, such processes are all that are available; but they are quite sufficient, for we do not want great exactness in the determination of χ^2. We merely desire to know whether the observations are with reasonable probability the result of a random sampling from the proposed theoretical distribution. Now let us examine the results. We see that for (2), (3–2), and (2–1) the normal curve is for practical purposes "impossible." As a matter of fact, we might have gone to the tenth figure without the probability being sensible in these cases. Further, (1) is highly improbable, the odds being about 1666 to 1 against its occurrence as a random sample. In the case of (1–3) the odds are 22 to 1 against a deviation as bad or worse than this, so that this is an improbable result. Lastly in case 3, and this only, we find the odds short, only 2·4 to 1, about, against it; a case such as this would occur on the average about twice in five trials. It is really the only case in which, under our present test, we could admit the normal curve.

Turning to the skew curve, we see that in three out of the six cases the odds are in its favour, namely, in (1), (3), and (1–3). It is not improbable in (2), the odds being only about 5 to 1 against it. It is improbable in (2–1) and very improbable in (3–2); in both of these cases, however, it is at least a million times as probable as the normal curve. Thus the skew curve is always markedly and often immensely superior as a method of describing the frequency to the normal curve.

Nor is it hard to discover grounds for its failure in cases (3–2) and (2–1), or for its lesser success in (2). The skew curve depending, as its constants do, on the fourth moment, takes much more account of outlying observations than the normal curve does. Let us consider how the χ^2 of these distributions is made up.

Absolute Equation (2). If the reader will look at the diagram (p. 294) of this distribution, he will observe the outlying observation on the left. There is a less marked one on the right. The skew curve endeavours, and fairly successfully endeavours, to account for these outlying observations by thinning its peak and stretching its tails—it thus becomes a much worse fit for the body of the observations than the normal curve. Beyond 3·5 on the left the skew curve leads us to expect ·3, about, of an observation, the normal curve only ·013 of an observation. Thus the outlying observation increases the χ^2 of the skew curve by about 3, but the χ^2 for the normal curve by about 73! In other words the outlying observation is not very probable from the standard of the skew curve; it is improbable enough to be considered practically impossible from the standpoint of the normal curve. If we reject this outlying observation as due to a momentary eccentricity of the observer, then with the *same* values of the constants of the curves the χ^2's are as 17 to 10, about, or the normal curve fits the body of the observations better than the skew curve. But this position is, of course, again entirely reversed if we fit the two curves afresh, re-calculating the constants without including the outlying observation in our data.

Relative Equation (3–2). A glance at the diagram shows the great irregularity of the distribution in this case. The outlying group on the left is here quite easily accounted for by the skew curve. It is immensely improbable on the basis of a normal distribution. The outlying group of three observations on the right contributes 17 to the χ^2 of the skew curve and only 2·7 to that of the normal curve. The peak costs the normal curve 29 and the skew curve 16. If we were to cut off the two extreme groups the χ^2 for the skew curve would be reduced to about 40, and for the normal curve, to about 75. Thus the skew curve, without re-calculation of constants, would still be immensely more probable than the normal curve. There is little doubt, however, that there is some source of change in the personal equation of Dr. LEE which has produced the anomalies in the relative judgment of Dr. MACDONELL and herself.

Relative Equation (2–1). The small probability of the skew curve and the "practical impossibility" of the normal curve depend entirely on the existence of the outlying observation to the right. The χ^2's in both cases would be reduced to about 24, and thus give probable results on the basis of random samplings if this outlying observation were removed. A re-calculation of constants would set the skew curve far above the normal, for its constants are more widely modified by outlying observations.

As I have already pointed out the value of χ^2 depends largely on where the range for the grouping of the frequencies is taken, and the tails largely determine what its value will be. But I have endeavoured to be equally fair to both theories, and rough as the numbers must necessarily be, we may still safely conclude that the skew curve gives infinitely more probable results than the normal. Indeed, with the rejection of an outlying observation or two, we could bring the whole skew-series into the range of probable random samplings, but we should fail to achieve this in the case of the normal curve without much "doctoring," which would have to be applied in certain cases to the very body of the observations and not only to its tails.

Personally while considering that the value of χ^2 is a very good criterion for the rejection or not of outlying observations, *as soon as a probable law for the distribution of errors has been determined*, I have thought it right not to reject one single observation* after the constants had once been determined, because I had in view the comparison of two different theories, and such rejection might apparently favour one or the other theory.

I now turn to the results for the bisection of lines; the probabilities for the random sampling in these series are given in Table XX.

* In the bright-line experiments 520 were originally made, as I supposed when we came to examine the recording strips some obvious slips or blunders would be found, and I left myself a margin of 20 for such. Only one experiment, however, No. 291, seemed to be a failure, the recording mark of one observer being in this case quite removed from the part of the scale occupied by the bright line.

TABLE XX.—Bisection of Lines.

		1.	2.	3.	3–2.	1–3.	2–1.
Skew curve	n'	16	21	16	17	21	23
	χ^2	11·36	17·27	14·90	8·84	20·28	23·99
	P	·7271	·6353	·4590	·9199	·4405	·3478
Normal curve	n'	17	20	17	17	22	23
	χ^2	13·30	22·04	20·31	9·34	21·38	25·65
	P	·6506	·2817	·2066	·8976	·4364	·2670

Now looked at from this standpoint, we see that not one of the distributions are improbable on the basis of either the skew or of the normal curve. The longest odds against a random sampling on the basis of a normal distribution are only 4 to 1, and on the basis of a skew distribution only 2 to 1. There is a clear and marked advantage in favour of the skew distribution, but it is nothing like so enormous as in the case of the bright-line series. If we take the distributions (1), (2), and (3) as giving independent probabilities of random sampling,* then the odds against these distributions as a result of random sampling from normal material are 24 to 1. Thus it seems that even in this case the normal law is somewhat improbable as a general law of distribution.

On the other hand, the combined odds against the system of distributions represented by the skew curves, are only 3·7 to 1; or looking at the problem rather differently: In the case of the normal distribution random sampling would give curves better than the observed in 20 per cent. of the trials, but in the case of skew distribution in only 5 per cent. of the trials. In other words, there is very great improvement in the closeness of fit produced by using skew distributions.

I place here the equations to the skew and normal distributions (a) and (b) respectively; remarking that the unit of y in either case is an observation, but the unit of x in the bright-line experiments is half a centimetre of the recording strip, and in the bisection experiments $\frac{1}{100}$ of the length of the line bisected.

* As we have found correlation in judgments, there is, of course, some assumption in this hypothesis but it will serve as a rough comparative test.

Bright-line Experiments.

Absolute Judgments.

(1.)

(*a.*) $x = 5·442,309 \tan \theta$.

$y = 92·5307 \cos^{8·386064} \theta \, e^{+1·078805\,\theta}$.

Origin of curve at $-$ ·38195.

(*b.*) $y = 87·060$ expt. $(- x^2/11·312,202)$.

Origin of curve at the mean, + ·07774.

(2.)

(*a.*) $x = 5·227,204 \tan \theta$.

$y = 100·806 \cos^{7·96724} \theta \, e^{·112,2496\,\theta}$.

Origin at $-$ 1·19399.

(*b.*) $y = 88·265$ expt. $(- x^2/11·005,434)$.

Origin at mean, $-$ 1·14483.

(3.)

(*a.*) $y = 53·359 \left(1 + \frac{x}{11·0856}\right)^{3·96821} \left(1 - \frac{x}{14·4504}\right)^{5·17262}$

Origin at the mode, $-$ ·59736.

(*b.*) $y = 56·837$ expt. $(- x^2/26·541,754)$.

Origin at the mean, $-$ ·44635.

Relative Judgments.

(3–2.)

(*a.*) $x = 11·17755 \tan \theta$.

$y = 21·2674 \cos^{15·34390} \theta \, e^{+5·89124\,\theta}$.

Origin at $-$ 1·78466.

(*b.*) $y = 59·537$ expt. $(- x^2/24·188,480)$.

Origin at the mean, + ·68275.

(1–3.)

(*a.*) $y = 48·890 \left(1 + \frac{x}{10·4055}\right)^{3·35160} \left(1 - \frac{x}{18·5911}\right)^{5·98815}$.

Origin at the mode, + ·2505.

(*b.*) $y = 51·580$ expt. $(- x^2/32·227,254)$.

Origin at the mean, + ·61145.

(2–1.)

(*a.*) $x = 8\cdot646433 \tan\theta.$

$y = 45\cdot7498 \cos^{10\cdot55017}\theta\, e^{-2\cdot976575\,\theta}.$

Origin at + ·28986.

(*b.*) $y = 62\cdot141$ expt. $(-x^2/22\cdot203{,}842).$

Origin at mean − 1·21518.

Bisection Experiments.

Absolute Equations.

(1.)

(*a.*) $x = 12\cdot89913 \tan\theta.$

$y = 60\cdot959 \cos^{31\cdot24937}\theta\, e^{-4\cdot441{,}508\,\theta}.$

Origin at + ·72873 hundredth of line.

(*b.*) $y = 81\cdot260$ expt. $(-x^2/12\cdot051{,}534).$

Origin at mean − 1·230 hundredths of line.

(2.)

(*a.*) $x = 48\cdot8323 \tan\theta.$

$y = 6\cdot038454 \cos^{261\cdot55}\theta\, e^{35\cdot61738\,\theta}.$

Origin at + 6·2061.

(*b.*) $y = 65\cdot072$ expt. $(-x^2/18\cdot793{,}284).$

Origin at mean + ·495 hundredth of line.

(3.)

(*a.*) $y = 71\cdot56246\left(1+\frac{x}{11\cdot349400}\right)^{5\cdot41665}\left(1+\frac{x}{6\cdot998{,}136}\right)^{3\cdot33995}$

Origin at mode + ·781519 hundredth of line.

(*b.*) $y = 75\cdot987$ expt. $(-x^2/13\cdot783{,}076).$

Origin at mean + ·377 hundredth of line.

Relative Equations.

(2–3.)

(*a.*) $y = 59\cdot43126\left(1+\frac{x}{16\cdot16672}\right)^{9\cdot763{,}885}\left(1-\frac{x}{13\cdot55284}\right)^{8\cdot185{,}180}$

Origin at the mode + ·2662 hundredth of line.

(*b.*) $y = 61\cdot633$ expt. $(-x^2/20\cdot949{,}076).$

Origin at mean + ·1230 hundredth of line.

(3–1.)

(*a.*) $y = 56\cdot5924\left(1+\frac{x}{16\cdot28385}\right)^{8\cdot216,093}\left(1-\frac{x}{12\cdot03998}\right)^{6\cdot074,832}$.

Origin at mode + 1·8495 hundredth of line.

(*b.*) $y = 59\cdot083$ expt. $(-x^2/22\cdot797,492)$.

Origin at mean + 1·5890 hundredths of line.

(1–2.)

(*a.*) $y = 56\cdot2136\left(1+\frac{x}{28\cdot15012}\right)^{35\cdot390655}\left(1-\frac{x}{37\cdot69023}\right)^{47\cdot384605}$.

Origin at mode − 1·8245 hundredths of line.

(*b.*) $y = 56\cdot688$ expt. $(-x^2/24\cdot763,446)$.

Origin at mean − 1·7120 hundredths of line.

Summing up the results of the above investigation as to random sampling, we conclude :—

(i.) That outlying observations render the skew curves a bad fit in one, and a very bad fit in a second case, and that in ten cases the observed results are very probable as random samplings from skew distributions.

(ii.) That the normal distribution is bad in one case and preposterously bad in four others ; it is probable in seven other cases, but in all cases less probable, and in five very much less probable, than the skew distribution.

We are thus led to much the same result as in our previous investigation of typical physical constants of the distribution, namely : that the axioms on which normality depends are not universally true, but that we require to use curves which will allow of a distinction between mode and mean, that will not assume an arbitrary relation between the fourth and second moments, yet which will pass gradually into the normal curve as we deal with material more and more nearly satisfying the fundamental axioms (α) (β) and (γ) (see pp. 274–275).

Such curves are supplied by the skew curves. If it be argued that these curves themselves involve relations between the first four and the higher moments, the answer is simply that we need only take such a number of independent moments that the bulk of frequency distributions can be represented as random samplings from our theoretical curves. It is idle to assert with LIPPS that if we have n frequency groups we must take n independent moments to describe the distribution, the *sine quâ non* of the problem is to describe with the fewest possible constants the distribution of a very great number of groups. Nor will arbitrary curves with six or seven constants do as well as a well-chosen curve with three or four.* The normal

* Tested in a variety of ways in a memoir on the general theory of curve fitting, which will shortly appear in 'Biometrika.'

curve in certain cases is a probable description, in a fair number of other cases it is a rough approximation, in many it is impossible. We must then start from its simple axioms (α), (β), (γ), and generalise in the next simplest manner. We assume that our contributory cause-groups are not indefinitely great in number, however numerous are the causes which determine the contribution of the group, that this contribution is not equally likely to be positive or negative, and finally that the contributions of the cause-groups are not independent but correlated quantities.* The simplest extension of the theory of GAUSS, LAPLACE, and POISSON in these directions leads us to the system of skew curves which have been applied in this memoir. I have treated them here purely from the experimental side. I have endeavoured to show in a fairly wide series of observations that the system of skew curves will, and the normal curve will not, satisfy the demands which we may fairly make on a theoretical frequency distribution. In another paper I shall consider the philosophical points which have been raised by EDGWORTH, LIPPS, and other recent writers. My present object has been to show certain failures in the ordinary theory of errors, and especially in its application to personal equation, and to show how existing theory may be widened so as to describe observations within the limits of the probable errors of the constants determined on the basis of random sampling.

12. *Summary and Conclusions.*

If we attempt to sum up the results reached, their importance seems to rest on the amount of weight that is given to the experimental material. Can we look upon this as typical of the measurements usually made by physicists and astronomers? I am unable myself to differentiate it, or to see causes for the high correlation of judgments which are peculiar to our experiments, and not to observations such as are daily made in the physical laboratory or the observatory. If this be so, then we must conclude as follows :—

(*a.*) The personal equation, while tending to a constant value, appears subject to fluctuations far exceeding those of random sampling.

(*b.*) These fluctuations in the case of two or more observers, whether dealing at the same time with the same phenomenon or measuring at different times the same

* Suppose, for example, that the cause-groups were those series of incalculable causes which determine (*a*) whether a coin shall fall head or tail uppermost; (*b*) whether an n-sided teetotum shall fall on one of p sides of one colour or not; (*c*) whether a card drawn out of a pack of np cards of p suits is of any particular suit. Then, if an indefinitely large number of coins be thrown together, the frequency distribution for heads satisfies all the fundamental axioms (α), (β), (γ) of the normal curve; if a finite number of teetotums be spun and the sides of the p-colour counted, we have satisfied (γ) only. If s cards be drawn simultaneously from our pack and the cards of one suit counted, then we have satisfied no one of the three fundamental axioms; there is correlation between the contributions of the cause-groups. This is only a rough illustration of the manner in which one or more of the fundamental axioms can be suspended artificially, but it is not without suggestion for the processes of nature.

physical quantity, appear to be "sympathetic." Thus there may arise a very considerable correlation of judgments between two observers assumed *à priori* to observe independently.

(*c.*) In addition to this psychological or organic correlation occurring in the case of absolute judgments, there is a spurious correlation which arises when two observers are referred to either a third observer as standard or to a common time or space element in each measurement as unit.

(*d.*) Errors of judgment whether relative or absolute far from universally exhibit the normal distribution of frequency. It is necessary to generalise this law of distribution, and this can only be done by supposing some or all of the axioms on which the normal law is based to be modified until we have a sufficiently general theoretical distribution, which will enable us to look upon the great bulk of observational errors as random samplings from the theoretical frequencies.

Even then we may expect occasionally outlying observations due to mistakes of record, or the interference of special causes of isolated occurrence, to render our distribution as a random sample improbable. But this raises the question of the rejection of improbable observations, which is common to any theory of distribution.

Practically it would seem :

(i.) That the correlation of judgments is a necessary factor in our appreciation of personal equation. The weight to be given to a combined observation, or to the combination of observations of two observers, depends upon a knowledge of this factor.

(ii.) That we should attempt not only to find the personal equation of two observers, but also the variations and correlation of their judgments. For this purpose it may be needful to make experiments *ad hoc*, mimicking the actual observations to be made as closely as possible, for there appears no method of determining these quantities from the relative as distinguished from the absolute judgments.

(iii.) That the existence of this correlation in judgments appears to vitiate very largely the existing theory of the probability of testimony ; that theory ought to be extended by the introduction of what we may term the psychological element ; an element which many may more or less unconsciously have found wanting, when they considered the weight which had to be given on the mathematical theory to the testimony of " independent " witnesses of the same series of events.*

(iv.) That great care should be used in applying the current theory of errors to observations until it has been shown that within the fluctuations of random sampling these observations really follow the normal law. If they do not, then the physical

* If Dr. LEE and Dr. MACDONELL assert that a bright line was in certain positions when the bell rang, their united testimony is very far from having the weight it would have on the old mathematical theory that they are independent witnesses, and yet they record perfectly " independently."

distinction between mean and mode,* the probabilities of negative and positive errors of the same magnitude being quite different, the abnormal concentration of errors round the mode, are all characters of the distribution, which must be taken into consideration, and which it is important to describe.

In concluding this paper I desire to heartily thank those who have aided me in its preparation. In the first place my gratitude is due to Dr. LEE, Mr. G. U. YULE, and Dr. MACDONELL for the time and care they gave in experiment and observation. In the next place I owe Dr. LEE special thanks for the constant assistance I have received in the laborious computations she has aided me in, and which are hardly obvious on the face of this paper. To Mr. K. TRESSLER I am indebted for great assistance in the conduct of the bright-line experiments, especially in the preliminary adjustments we had to go through before we got our apparatus into efficient working order. He has also prepared from the calculations of Dr. LEE and myself the whole of the frequency diagrams. The work of experiment and reduction has extended over nearly six years, during which considerable progress has been made (*e.g.*, by Mr. SHEPPARD's discovery of the best corrective terms for the moments) in statistical theory, and thus all our data have not been dealt with in an absolutely uniform manner;† but the divergences due to method are small as compared with the probable errors, and we have taken great care by duplication of calculations to avoid as far as possible arithmetical blunders.

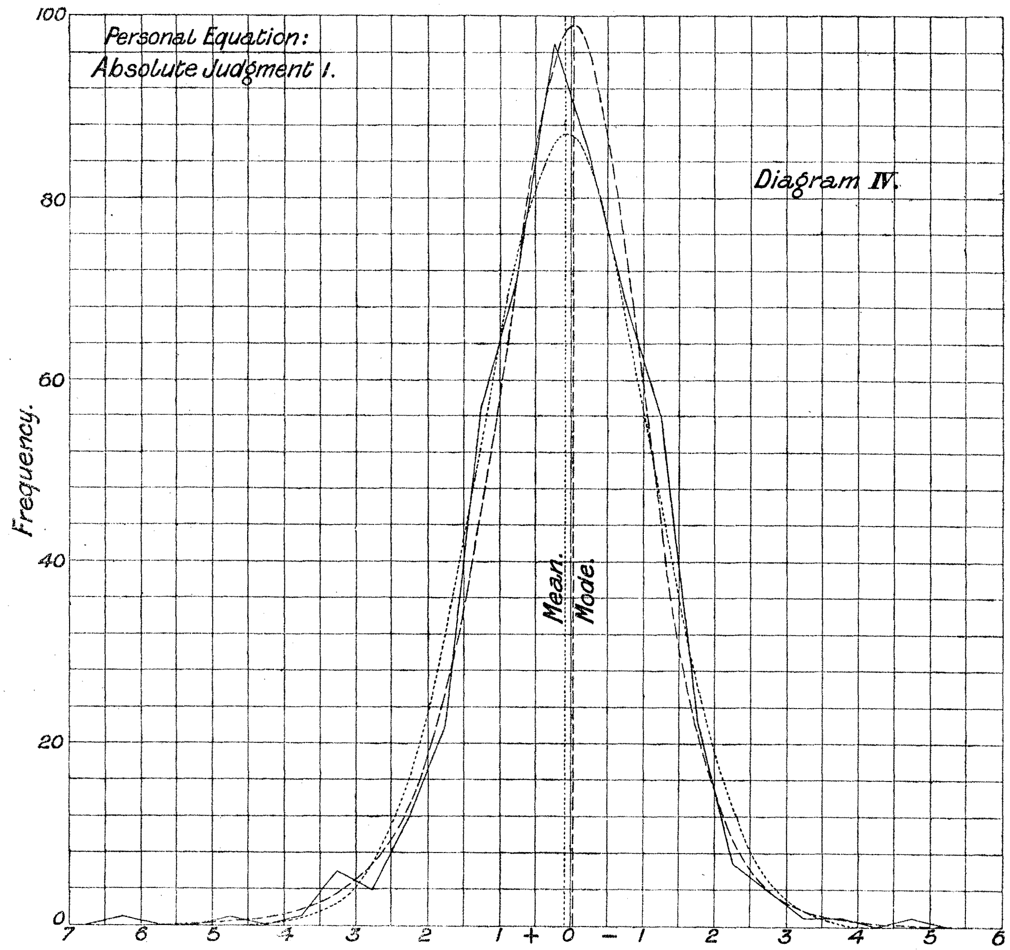
* Some American writers persist in taking the maximum group of observed frequency as the mode. But the fluctuations of random sampling make such a determination of the mode in many cases quite futile: see for examples my Diagrams VI., IX., and XII. The mode is where dy/dx vanishes for the theoretical frequency curve, and is not visible on mere inspection of the observations.

† The calculations for the bisection of lines were in part made on the grouped observations without SHEPPARD'S corrections, *i.e.*, with the value of the mean error as given in the usual theory.

MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

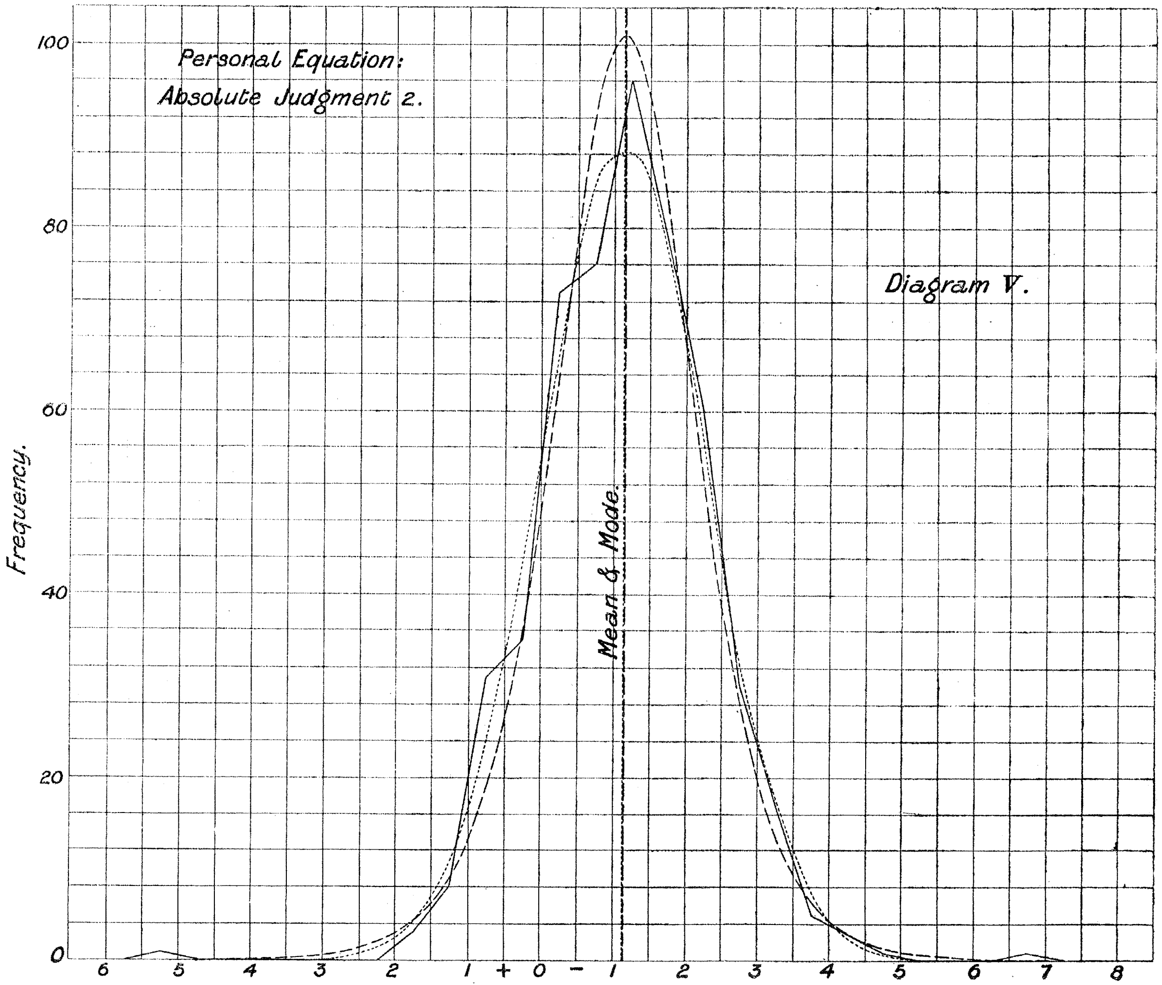
PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A

Bright Line Experiments.

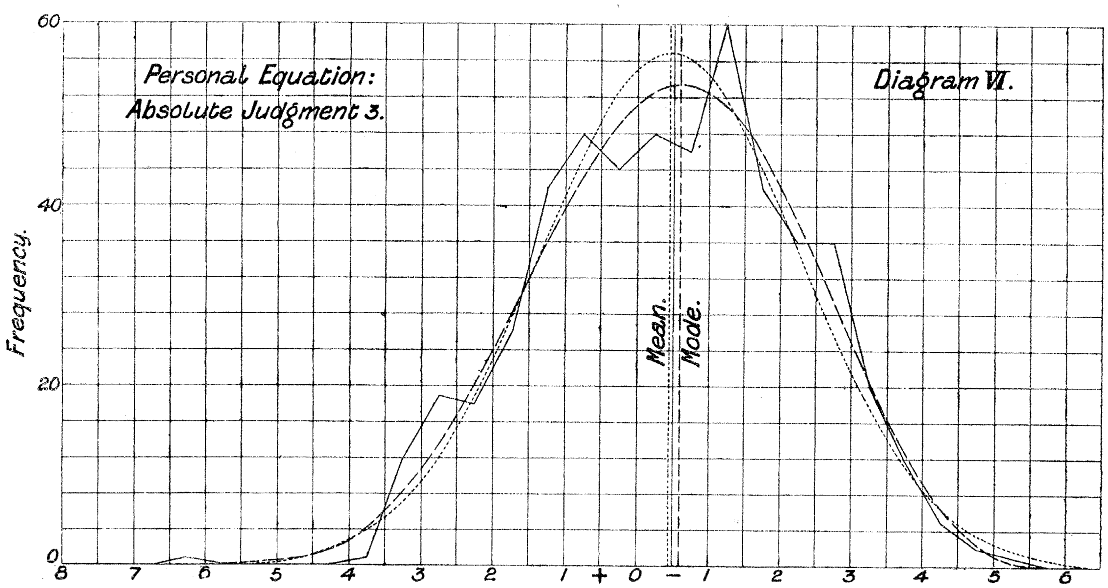


PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Bright Line Experiments.

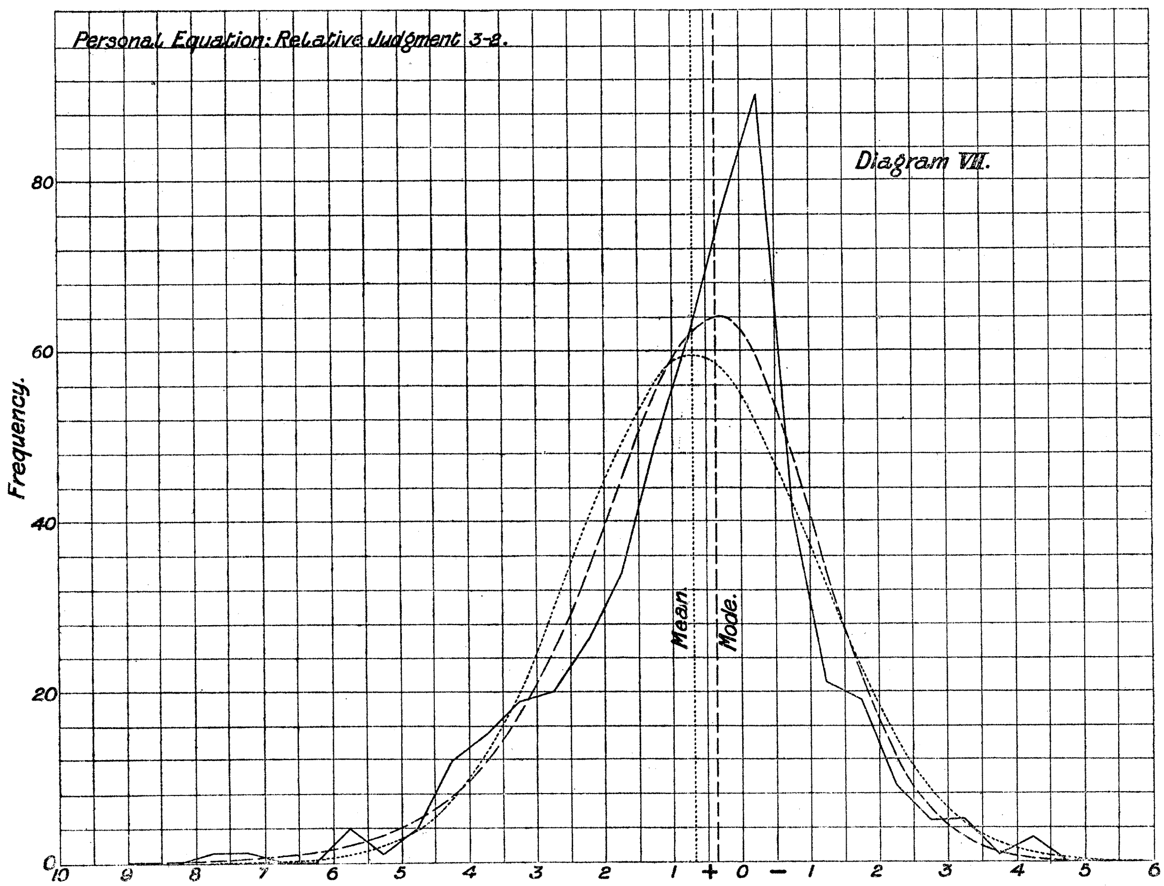


Bright Line Experiments.

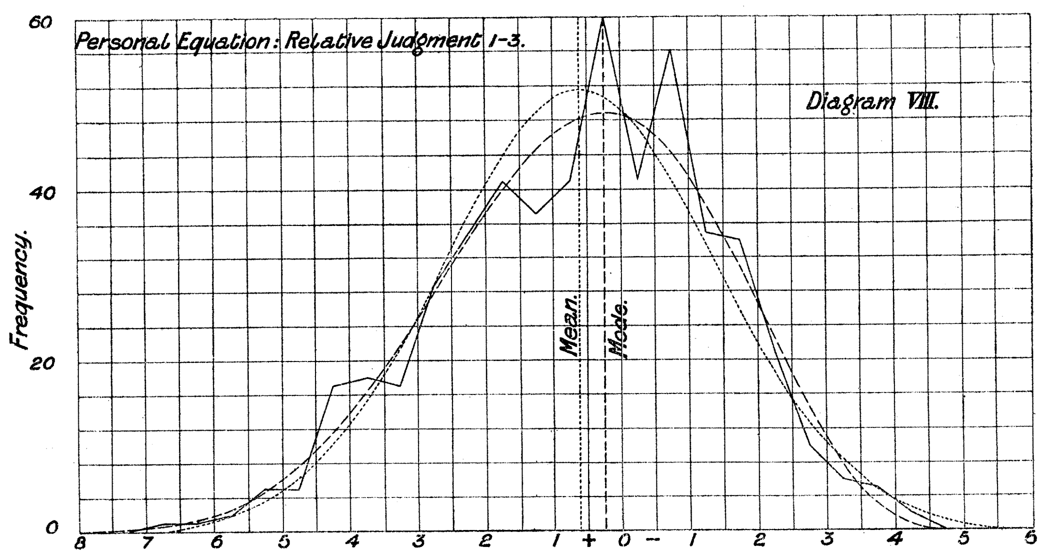


PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Bright Line Experiments.

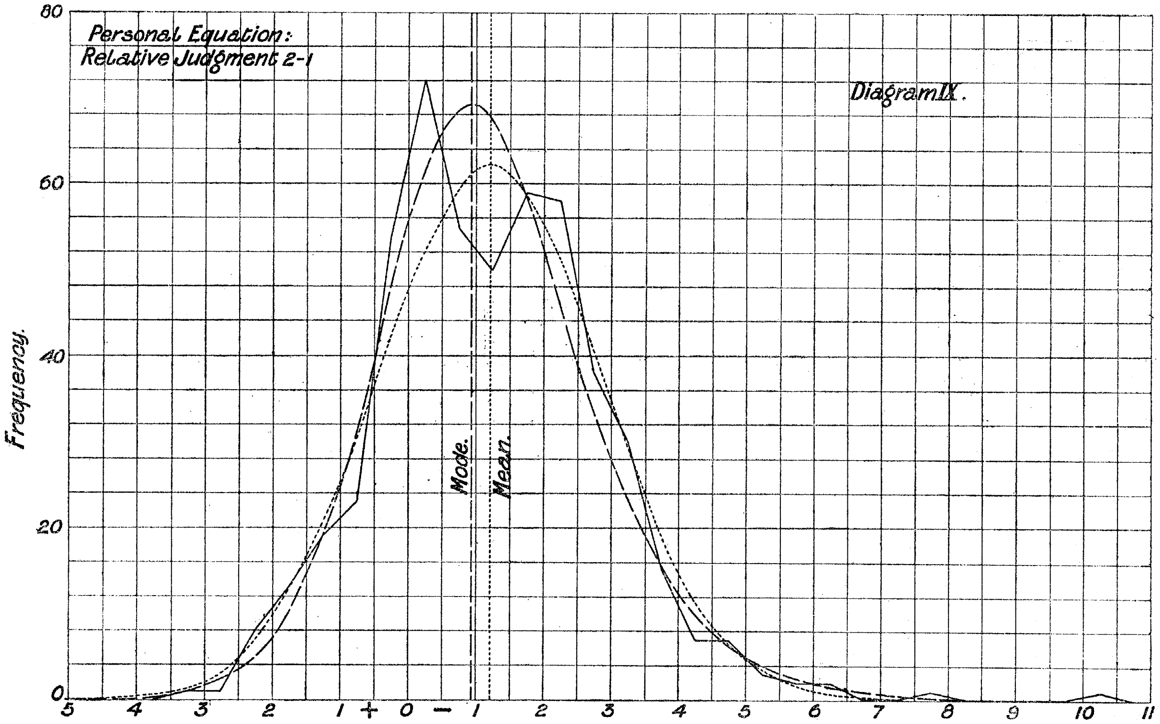


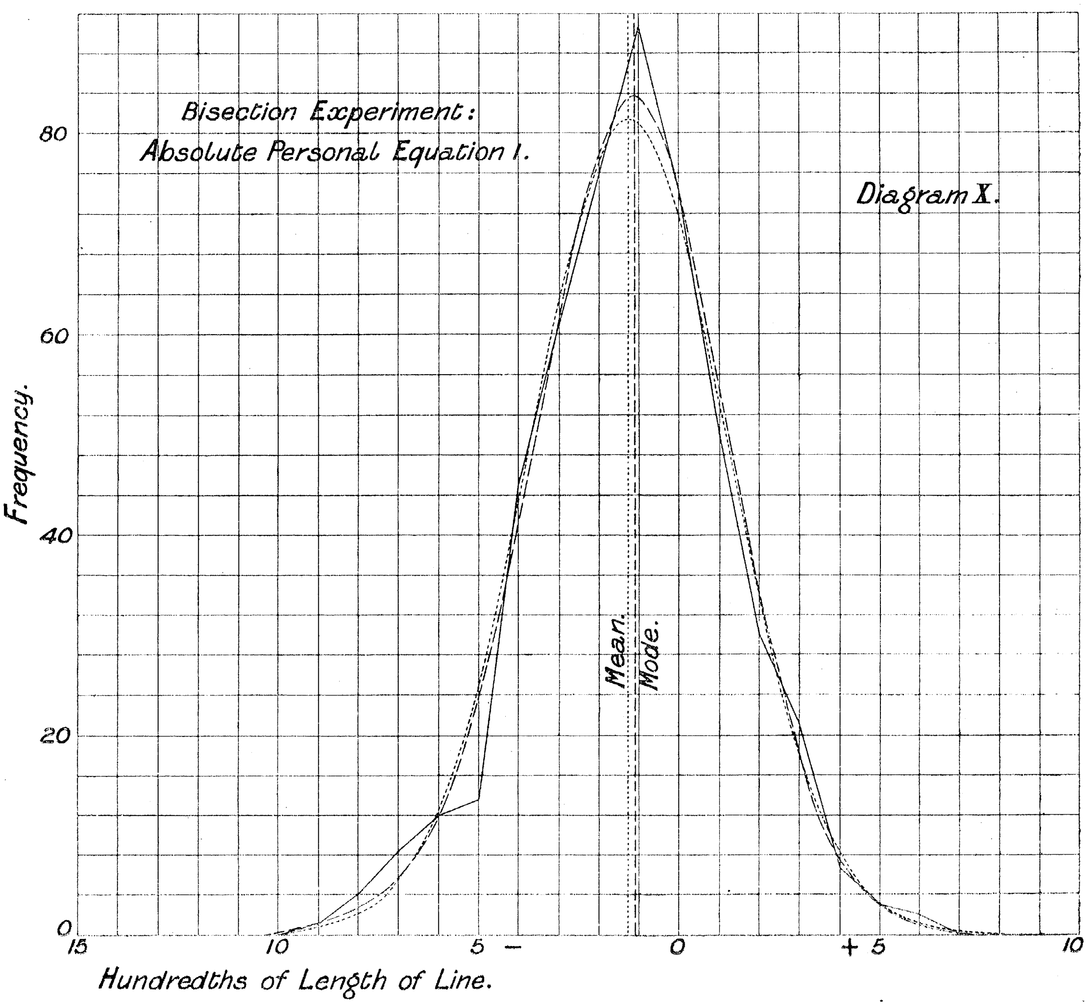
Bright Line Experiments.

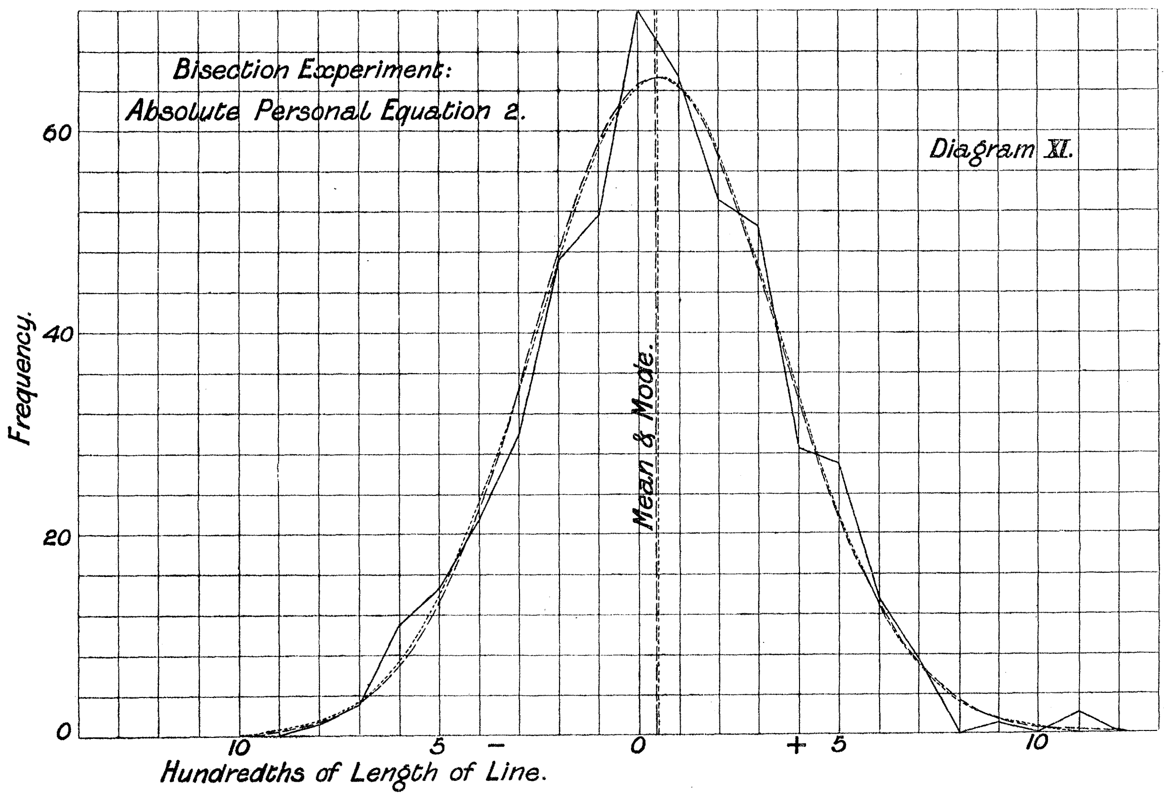


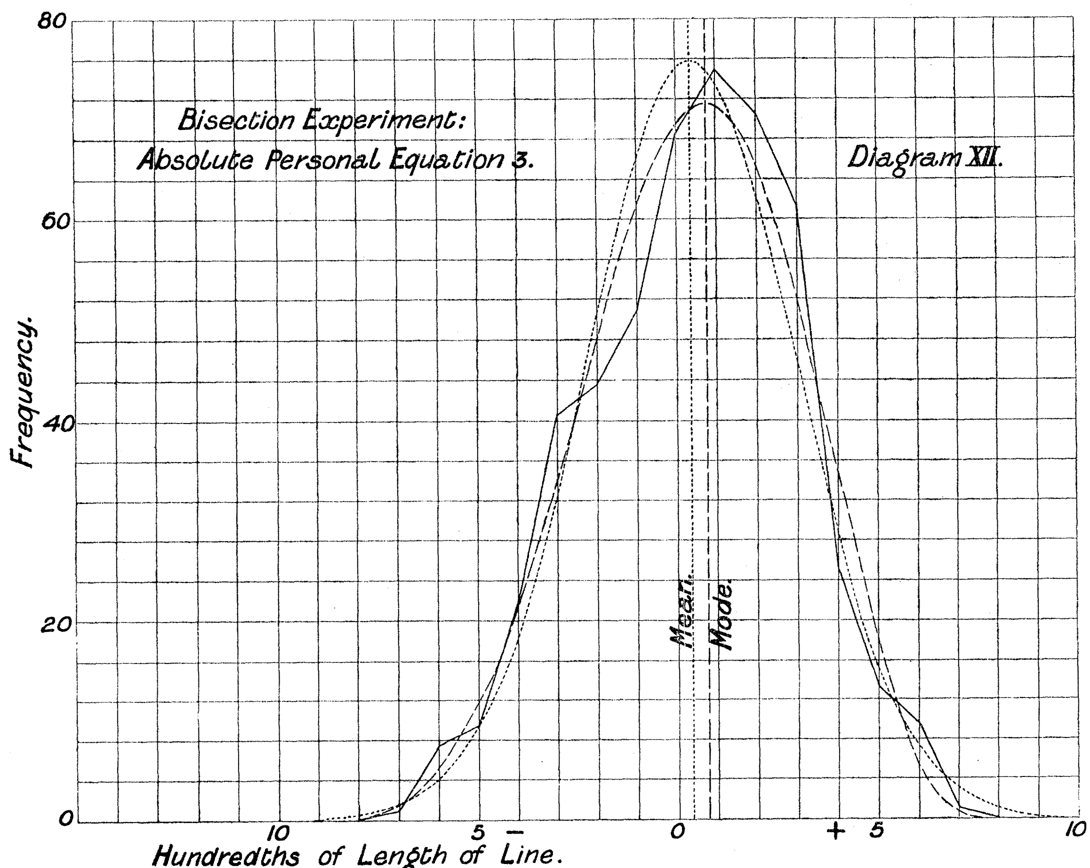
PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Bright Line Experiments.

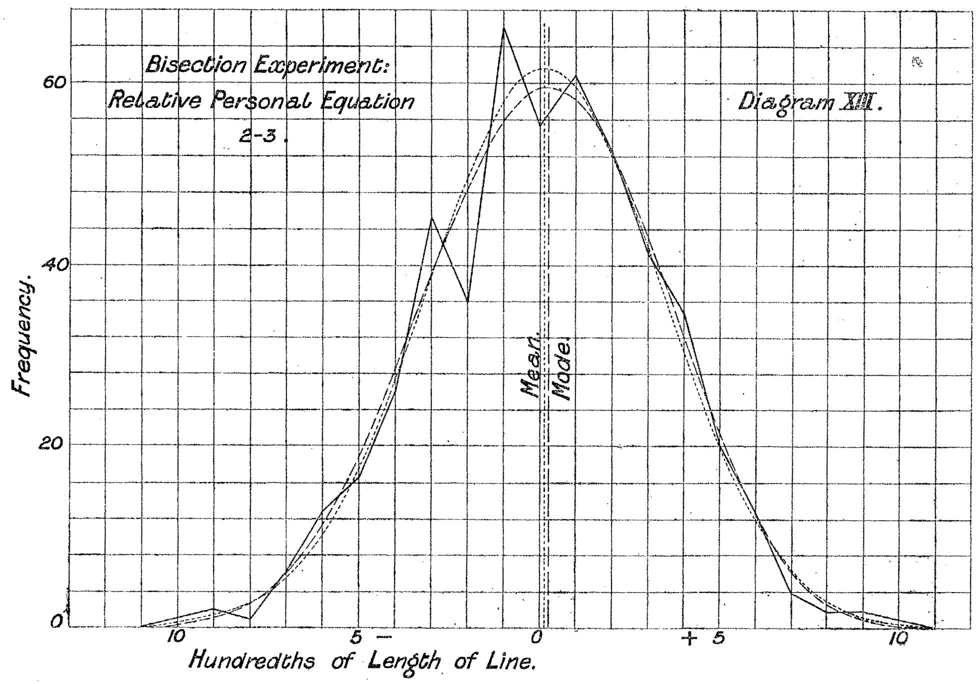


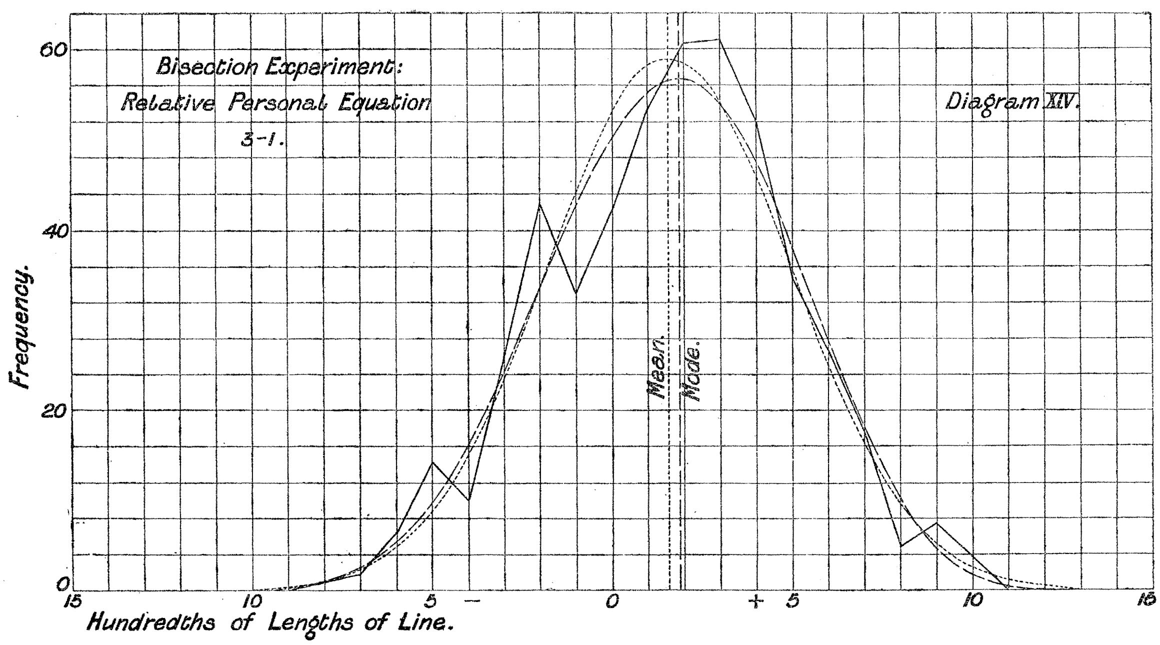




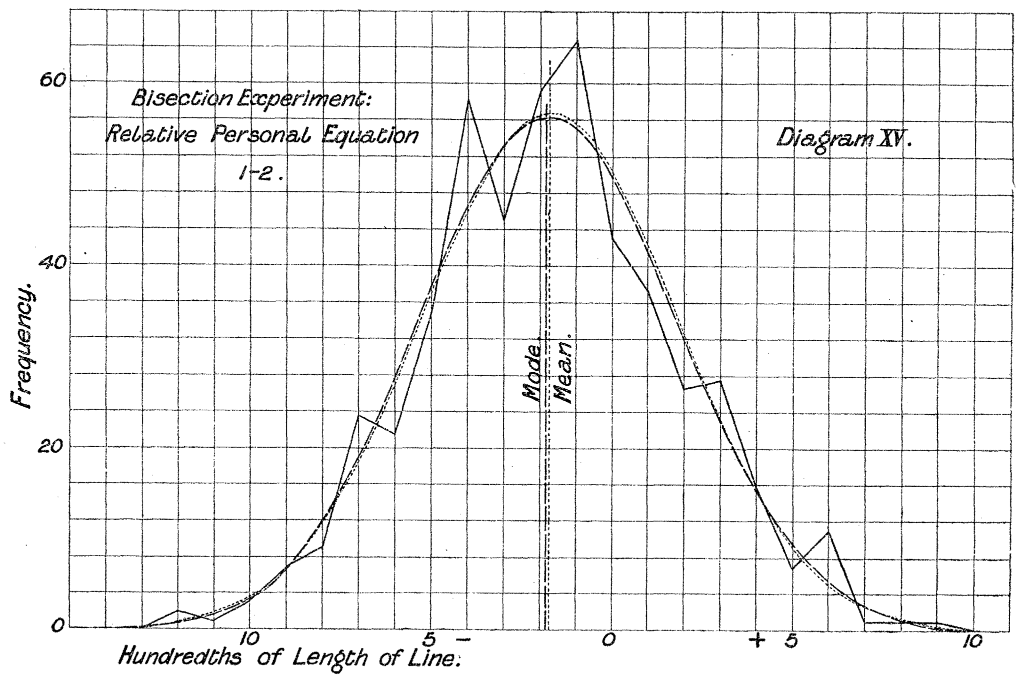


PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES





PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES



PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES