THE LOGIC OF DISCOVERY

by

W. A. Shewhart

*Paper presented before the Physics Colloquium, March 30, 1931.

INTRODUCTION

Logicians and philosophers have always talked a great deal and written many books about the logic of discovery or the theory of induction. Nevertheless it sometimes seems that what Mark Twain said about the weather, namely, that it is something that everyone talks about but no one does anything about, might be applied to the logic of discovery so far as scientists are concerned. In the light of this situation there is every reason to ask the question: Why should an engineer think about the logic of discovery? In what follows, I shall try to answer this question. I shall try to show why it is that some of the recent developments in the theory of rational induction give promise of getting engineers out of many of the pitfalls of loose reasoning into which they have fallen and of making it possible for the first time to establish ways and means of securing economic control of quality of manufactured product.

To clear away any doubt about the practical importance of this question, let us consider a few specific instances before entering upon the general discussion. Let us start with the problem of sampling involved in the inspection of the quality of materials. It might be thought that the best way of avoiding the sampling problem is to make a one hundred per cent inspection. However, there are two major reasons why this cannot be done. In the first place, it is not economical and in the second place, a one hundred per cent inspection cannot be given where the test is destructive.

As an example of the latter case, we as a company buy approximately two million dollars worth of steel strand per year, one of the most important properties of which is tensile strength. Such strand often comes in lengths of 5,000 ft. wound on a single reel. The practice in such a case is to cut a 24" piece from the exposed end of the reel and to make a tensile strength test on this piece. A question that the inspector must answer is: What does this single value tell about the tensile strength of the 5,000 ft. of strand?

Now, how does the latest development in the logic of discovery influence the answer to this question? It shows something very important: it shows that only under certain conditions does a sample give any appreciable assurance about the quality of material from which the sample was drawn. These are the conditions of economic control in the technical sense developed in these Laboratories.

We may take as another problem typical of those of interest in inspection engineering that considered at a recent symposium on welding[1]. In speaking of the Charpy impact test results, one author stated that this test indicates the ability of engineering materials to absorb energy quickly and may be used as an indicator of toughness. He also stated

---

1. Held at Pittsburgh, March 18, 1931, under the auspices of the A.S.T.M.

that the ductility of the material indicates toughness. He then showed the results given in Fig. 1 to indicate as he said the correlation between impact value and the static ductility values. He commented that the correlation is not very good. The real question is: Is there <u>any</u> correlation, and if so how can we measure it? We certainly cannot give a rational answer to this question without making use of recent developments in the distribution theory of measures of correlation.

Fig. 1

It is not alone the poor inspection engineer, however, that has his worries of this kind. If we attend any scientific or engineering meeting, we hear papers on the measurement of this and on the measurement of that, and other papers on the discovery and measurement of relationship of one thing to another. All of us have been present at such meetings when several in the audience failed to be convinced that the reported measurement was a satisfactory estimate of the objective thing to be measured and that the reported relationship was the true or objective relationship that was sought.

So often under these conditions we hear the argument that although a conclusion may appear theoretically sound, nevertheless it is not consistent with common sense. In fact you will find it interesting to note in articles and books on engineering, physics, and science in general, the number of times an appeal is made in just this way to what is called common sense. Now, if there is anything that can be stated with great assurance about common sense, it is that it is not common to any two individuals. If we as engineers and scientists wish to justify our conclusions on a rational basis, we must meet the following question squarely: What scientific basis have we for measuring the faith that we should have in any announced discovery in engineering or science?

To keep our feet on the ground, let us look at Fig. 2 taken from an article[1] in the current issue of Metals and Alloys. What faith are we justified in placing in the various relationships there shown? It is difficult to answer without knowing the dispersion of the observed values about the indicated points. Could such results be used justifiably as a basis of economic

- -

1. Thompson, J. G., "Bismuth Alloys", February, 1931.

Fig. 2

control of Bismuth alloys? The answer to the last question is: No, for reasons that will be given later.

As another illustration, how are we to interpret Fig. 3 reproduced from a recent book[1] on materials. Are we to judge from the title of this figure that the tensile strength of wrought iron is <u>functionally</u> related to the temperature of the iron? If we do, we are being misled because as we shall soon see the available data do not satisfy the conditions requisite for our belief in the existence of such a relationship.

Beginning with the work of John Stuart Mill[2] - in particular his five cannons of inductive inference - much has been done in an attempt to develop a <u>formal scientific</u>

- -

1. Judge, A. W., <u>Engineering Materials</u>, Vol. III, Sir Isaac Pitman & Sons, Ltd., 1930.
2. Mill, John Stuart, <u>System of Logic</u>.

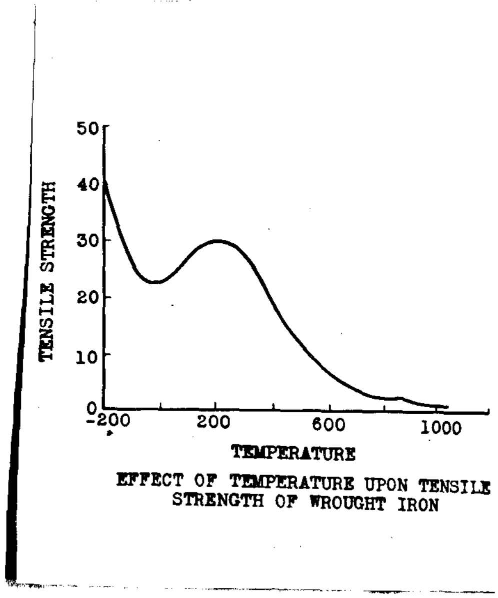

Fig. 3

method of induction or discovery[1].

One of the reasons for considering anew the method of induction is that within the past few years we have witnessed the rapid growth of the statistical concept of physical properties and physical laws. Whereas within our memories, physical properties were thought of as constants, and physical laws, as mathematical or functional relationships, we are now living in an age when physical properties and physical laws are thought of

---

1. Within just the past two decades numerous articles and many books have appeared. Some of the more important contributors in this period have been Bertrand Russell, A. N. Whitehead, Jean Nicod, J. M. Keynes, C. D. Broad, H. H. Dubs, R. J. Carmichael, Sir James Jeans, A. S. Eddington, Harold Jeffries, W. E. Johnson, A. D. Ritchie, J. Neyman, R. A. Fisher, and Egon Pearson.

as being inherently statistical in nature.

With the introduction of this new concept comes the necessity for a modification of the concepts of causation of the older logic of discovery in which the canons of Mill and modifications thereof play such an important rôle. We must come to appreciate the indeterminateness of causation and the significance thereof particularly in the economic production of great quantities of the same kind of material or great numbers of the same kind of apparatus made in accord with given specifications.

It is perhaps trite to say that without mind there would be no discovery. However, this human element depends upon two things: native intelligence[1] and environment or training. It is therefore of interest to know how the controllable part of the human element can best be trained for discovery.

In proceeding we shall limit ourselves to a consideration of two kinds of discovery; viz., constants of nature and relationships or physical laws.

TWO GENERAL METHODS OF DISCOVERY

Let us consider the simplest kind of problem in discovery; viz., that of discovering[2] an objective fixed magnitude such, for example, as the charge on an electron. Suppose we agree on a method of measurement subject, of course, to errors

- -

1. Of course, psychologists are still arguing as to what they mean by intelligence.
2. We usually say measuring.

of measurement. As an estimate of the charge, we take the average of several measurements. The question immediately arises, however, as to how many measurements we should take. This necessitates some assumption as to the method of approach of the observed average to the objective magnitude as we increase the number of observations.

It is customary to assume in such a case as a fundamental postulate that, if we keep all conditions essentially the same, the average $\bar{X}$ of n observed values X_1, X_2, ..., X_i, ..., X_n, approaches the true objective value $\bar{X}'$ as a statistical limit L_s expressed formally as follows:

$$\underset{n \to \infty}{L_s} \bar{X} = \bar{X}'. \qquad (1)$$

The significant characteristics of such a limit are considered in some detail in a forthcoming publication[1]. All that we shall do here is to give an illustrative example making use of the observed values of the charge on an electron as given by Millikan[2]. The successive points in Fig. 4 give the averages of 1, 2, 3, 4, .., n of Millikan's observed values of the charge on an electron. Note that as we proceed to the right

- -

1. Shewhart, W. A., Economic Control of Quality of Manufactured Product, to be published by D. Van Nostrand Company.
2. Millikan, R. A., The Electron, University of Chicago Press, 1917.

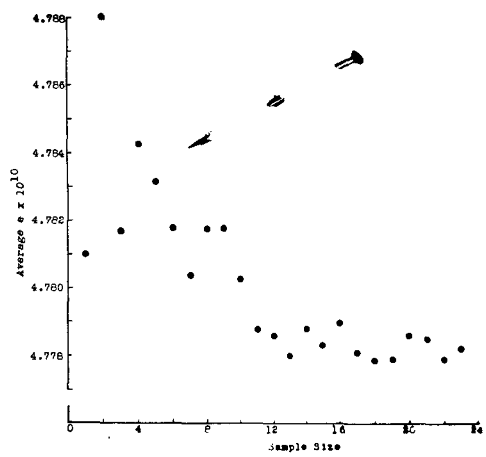

Fig. 4

the observed oscillation in the points rapidly tends to decrease, and as we say, the average approaches a statistical limit characteristically different, of course from a mathematical one. If the observed average approaches the true value in accord with (1), we can say that, in general, the more observations we take the more likely it becomes that our observed average differs by less than some preassigned value ε from the true value $\overline{X}'$.

In other words, under these assumptions, we can get a relatively good picture of the increase in information accruing through an increase in the number of observations. It is true, of course, that this estimate rests upon still further assumptions as to the nature of the distribution of errors of measurement. Without any other assumption, however, than that the statistical limit (1) applies, we may say that the standard deviation of the average decreases inversely as the square root of the number of observations. The important point in this connection is that the significance of increase in sample size n depends upon one little word <u>if</u>, in the sense that increasing the size n of sample assures us in the statistical sense that we will get closer and closer to the true value $\bar{X}'$ as n is increased <u>if</u> the same essential conditions are maintained so that the average $\bar{X}$ satisfies the condition of a statistical limit and <u>if</u> the statistical limit is the true value $\bar{X}'$. In other words, we may say that, if there is no constant or systematic error, we can get some indication of the significance of increase in sample size.

At this point, the important question naturally arises: How do we know that we have succeeded in eliminating the constant and systematic errors? The answer is that we <u>never know</u>. In other words, the formal machinery of the scientific method for this case, resting as it does upon the assumption of the existence of a statistical limit, fails to give us assurance that we are getting closer and closer to the truth unless we start with certain kind of data.

To get good data the human element must enter. Scientific method offers no formal program for getting the proper kind of data. In the last analysis, even in this simple case, the results of the application of scientific method are personal in character.

What is the practical significance of this conclusion insofar as it relates to the large quantities of data accumulated in the production and inspection of the quality of manufactured product? It is simply this: Unless the data are good in the sense that increasing their number leads closer and closer to the objective reality, analysis of the data can yield little of value.

In the research laboratory emphasis is usually laid upon the quality of the data - not upon the number. In engineering, however, the question most often asked has been: How many observations shall be taken? This is indicative of the misplaced emphasis on numbers of data irrespective of whether or not the data are good. It was this situation in respect to inspection data that gave cause for the recent developments of criteria to indicate whether or not the data are good.

Briefly such a criterion essentially consists in making use of modern statistical distribution theory in the establishment of limits within which observed variations in good data may be expected to fall. Unless a set of data exhibit technical control in this sense they cannot be used as a basis of prediction. Without the use of such criteria the results of sampling are, in general, non-intelligible. Thus we see the importance of some of the recent developments in the logic of discovery.

Let us pass now to a consideration of the other general method of increasing assurance that a result of measurement is close to the true objective value. We shall use as an illustration the measurements of the charge on an electron by different methods. Fundamentally a given method corresponds to linking of observed facts through assumed relationships in such a way that they give an indirect measure of the charge. Schematically this may be represented as in Fig. 5. The crosses in the figure represent observed facts. The curves indicate theoretical relationships involving certain facts and the charge

Fig. 5

on an electron represented schematically as the point to which all of the lines converge. The more relationships of this character which give approximately the same value as an estimate of the charge on an electron, the more faith we have in our resultant measure.

It remains, however, to estimate the increase in assurance with increase in the number of different ways of measuring the charge. There is today a tendency to appeal to probability theory and to state that one result is probably true or more probable than another. For example, Millikan[1] in a recent paper uses in the title the phrase "the most probable". One of the important points that will be made as we proceed is that scientific method does not provide any means of finding the most probable value of a measurement in the objective sense. Even in the simple case of repetition of measurement, it does not provide a probability that is independent of an assumption explicitly stated or implied; it provides no means whatsoever of giving a probability where various indirect methods of measurement have been employed.

Hence it is that, although the second process of discovery unquestionably gives greater assurance than the first in the sense that agreement of let us say 25 measurements

---

1. Millikan, R. A., "The Most Probable 1930 Values of the Electron and Correlated Constants", Physical Review, Vol. 35, May 15, 1930, pp. 1231-1237.

divided equally between 5 different methods of measurement gives greater assurance than 25 measurements by a single method, nevertheless we have no formal method of estimating the increase in assurance given by increasing the number of methods of measurement.

NATURE OF RELATIONSHIP

A. Rules

A low order of relationship is that expressed in the rules of every day experience such as: robins go south in the winter and return in the spring; a furrowed brow indicates deep concentration; the good die young, and statisticians are mostly liars. Such rules are often only partly true - a furrowed forehead may indicate concentration and it may indicate a stomach ache. Most of us would not agree that all of the good die young and at least a few of us will not agree that statisticians are mostly liars.

B. Functional Relationship

One of the first results of the development of the scientific method was the introduction of the concept of an important kind of relationship; namely, functional. For example, measurements of pressure, volume, and temperature of gases led to the conception of an ideal gas in which these three properties are related functionally one to another.

Such experimental results have led to the conception of objective functional relationships existing between the macroscopic properties of physical things expressable in the form

$$f(X_1, X_2, .., X_i, .., X_m) = 0. \qquad (2)$$

In most instances the function f is assumed to be continuous over the assumed range of relationship, and the m different properties represented by the X's are assumed to be such that if m - 1 of them are fixed, the values that the other one may take are determined, the X's being assumed to be mathematically independent. A simple example of this kind is the equation of state of a perfect gas.

C. Statistical Relationship

Two variables are said to be statistically related under the conditions:

(a) That the observed fraction p of the number of times that simultaneously observed pairs of values of these two variables fall within a given area in the plane representing these two variables, approaches as a statistical limit L_s some fixed value p', as the number n of observations approaches infinity, or in other words that

$$\underset{n \to \infty}{L_s}\ p = p'. \qquad (3)$$

(b) That the frequency distributions in the array of one variable are not the same for all of the arrays.

We may represent this kind of relationship formally as follows:

$$f_s(X_1, X_2, \ldots, X_i, \ldots, X_m) = 0, \tag{4}$$

where the significance of f_s is that just indicated, being, of course, different from the f in (2). The force of condition (a) is to make constant the probability[1] that an observed set of values of X will fall within a given element of volume dX_1, dX_2, .., dX_i, .., dX_m, or, in other words, to make the system statistically specified in terms of the m different characteristics. Whereas in the functional relationship the X's are assumed to be independent quantities, this condition is not satisfied in the statistical relationship, it being assumed that the various pairs of values are correlated.

A typical observed set of data indicating the presence of a statistical relationship between two variables is that given in Fig. 6 showing the scatter diagram of modulus of rupture

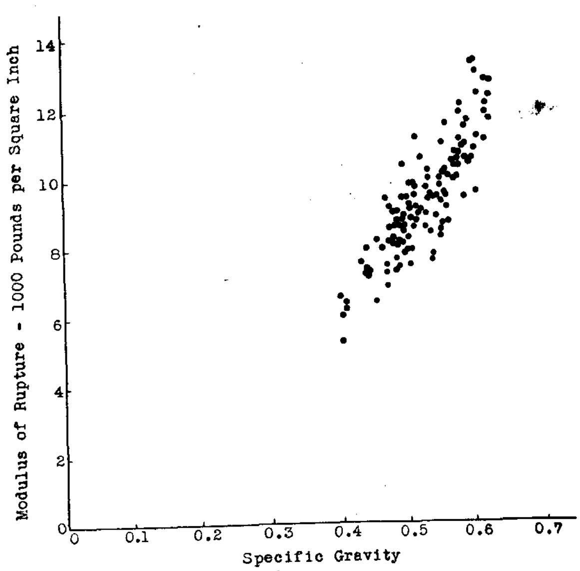

Fig. 6

1. Considered as a statistical limit.

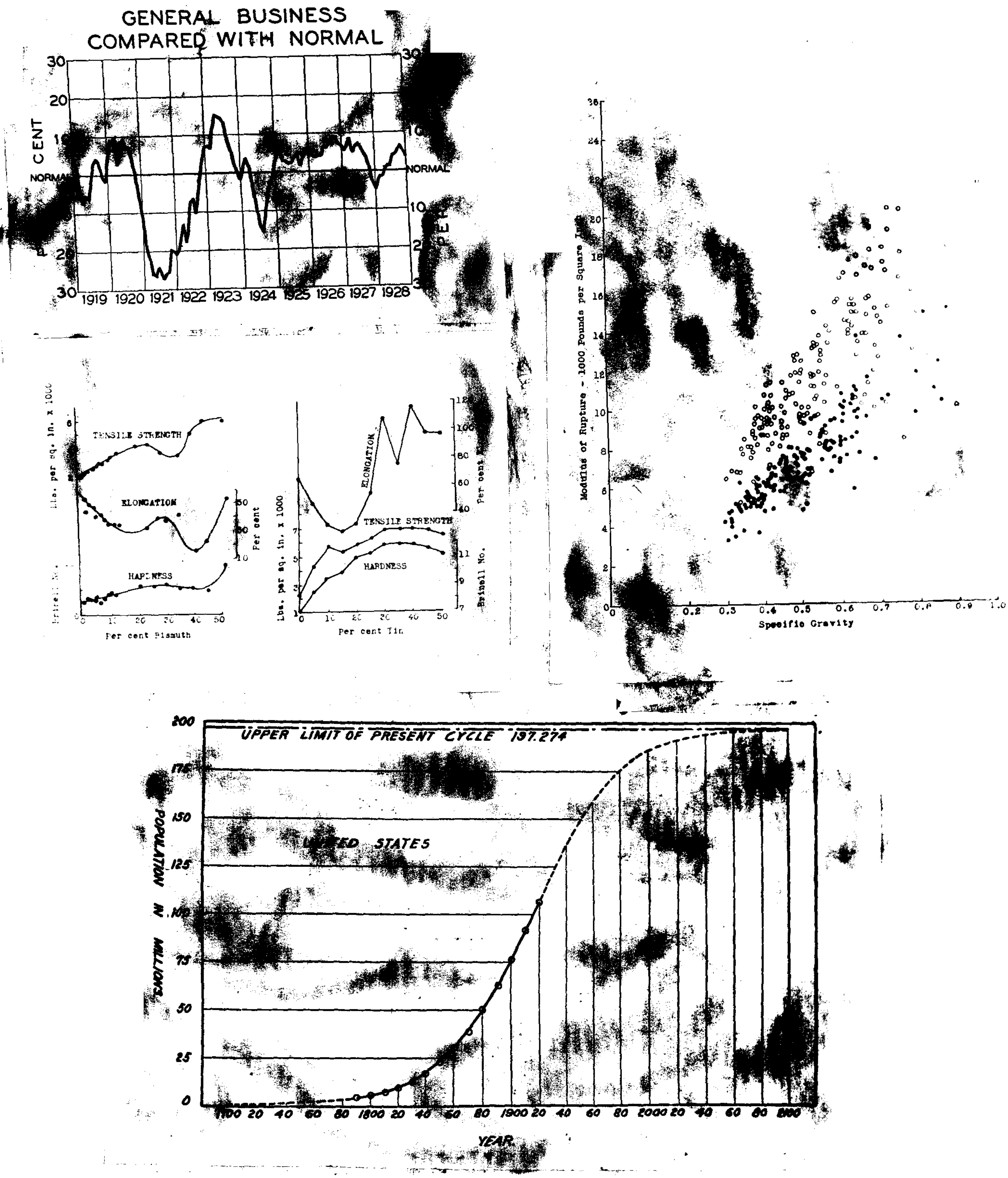
GENERAL BUSINESS
COMPARED WITH NORMAL
NORMAL
1919 1920 1921 1922 1923 1924 1925 1926 1927 1928
TENSILE STRENGTH
ELONGATION
HARDNESS
Per cent Bismuth
Per cent
Lbs. per sq. in. x 1000
Per cent Tin
Brinell No.
Modulus of Rupture - 1000 Pounds per Square In.
Specific Gravity
UPPER LIMIT OF PRESENT CYCLE 197.274
UNITED STATES
POPULATION IN MILLIONS
YEAR

Fig. 7

In the general case, the first step is to discover the set of m variables $X_1, X_2, \ldots, X_i, \ldots, X_m$ that specify the thing under consideration. The practical significance of this general problem is apparent when we note that the complete specification of the quality of a given kind of raw material or finished piece of product inherently involves the problem of discovering and stating some equation of state to be used as a standard. In the present state of our knowledge, we are not in a position to do more perhaps than set up conditions of statistical control of certain quality characteristics.

A. Functional Relationship

Let us assume that two variables X and Y satisfy the functional relationship

$$Y = f'(X, \lambda'_1, \lambda'_2, \ldots, \lambda'_i, \ldots, \lambda'_{s_1}), \tag{5}$$

where f' is the unknown true functional form and the λ''s are the unknown parameters to be discovered.

The problem of discovery may be divided into three steps:

(a) Obtain a set of n pairs of values

$$Y_1, Y_2, \ldots, Y_i, \ldots, Y_n$$

$$X_1, X_2, \ldots, X_i, \ldots, X_n$$

from which to discover (5).

(b) Discover the functional form f'.

(c) Discover the values of the s_1 parameters.

Let us consider these three steps in the above order.

Good Data

It goes without saying that the n pairs of values should be free from constant and systematic errors of measurement. Far more important is the fact that all variables other than X which affect Y should either be eliminated or kept constant.

The important question is: What are the factors that must be controlled and how can we be sure that they have been controlled in getting an observed set of n pairs of values of X and Y? Obviously we cannot be sure of our answer to this question. Success depends upon our previous experience and our knowledge of the details of the problem in hand. Furthermore, it depends upon our intuitive, speculative, and postulational faculties. To a man of long experience in research, this statement is trite. I fear, however, that when we come to consider the great quantities of data representing measurements of the quality of manufactured product all the way from raw material to finished product, we shall find that the importance of this characteristic of good data is often overlooked. In fact, I have been told by engineers that by taking great masses of data, errors and mistakes cancel out. The fact is, of course, that quantity of data in no way compensates for quality - errors and mistakes do not, in general, cancel out. One of the crying needs in many phases of engineering work is better data and this often means that greater emphasis needs to be placed upon the human element and less emphasis upon the numbers of routine data.

Functional Form f'

Since we never know f', the best that we can do is to choose some functional relationship involving s parameters such as

$$Y = f(X, \lambda_1, \lambda_2, \ldots, \lambda_i, \ldots, \lambda_s). \tag{6}$$

Needless to say, the function f may not be and perhaps is not the true objective f'. Furthermore, it is not likely that the assumed functional relationship will involve the same number of parameters as the objective relationship. In other words, it is likely that $s' \neq s$. For example, it is customarily assumed that the best known laws of physics are only approximations to the true laws.

The choice of the functional form obviously depends upon the investigator and upon his knowledge of the field in which he is working. If in an entirely new field, he will not likely have any a priori basis for choosing a functional form before taking the data. If, however, he is working in a familiar field, he may from a study of available facts and theory in this field be led theoretically to try a given functional form.

Needless to say, it is possible to find a functional form involving as many parameters as there are observed pairs of values such that the function will pass through every observed point[1]. There is no reason, however, to believe that

---

1. At least where there is only one observed value of Y for a given value of X.

such a function has much physical significance, if for no other reason than that it is unlikely that the observed points lie on the curve of objective relationship.

The chosen functional form may be of a closed type involving a fixed number of parameters or it may be of the open type such as a series of orthogonal functions wherein the number of parameters is not fixed. In the latter case there is involved an additional choice as to the number of parameters that it is worthwhile trying to discover.

Determination of Parameters

Having decided upon the functional form f and the number s of parameters to be used, the next step is to choose the best values for the estimates of parameters. But what shall we take as the best values? This is an open question.

Several formal methods are available for finding a set of s parameters such, for example, as the determination of their values from s observed pairs of values of X and Y; the graphical method; the method of zero sum proposed by Norman Campbell; and perhaps the most widely used method, that of least squares which states that the sum

$$\sum_{j=1}^{n} \left[Y_j - f(X_j, \lambda_1, \lambda_2, \ldots, \lambda_i, \ldots, \lambda_s) \right]^2 \qquad (7)$$

shall be made a minimum.

In a given problem, these formal methods lead to different sets of values as estimates of the parameters and we are faced with the problem of determining which one of the sets is the best one in a given case. In the absence of a universally accepted criterion, the choice of the method of determining the parameters becomes a matter of opinion.

As an indication of the wide divergence of opinion about the method to be used, let me quote first what E. T. Whittaker and G. Robinson say, in general, about the graphical method in the preface to their book The Calculus of Observations: "When the Edinburgh Laboratory was established in 1913, a trial was made, as far as possible, of every method which had been proposed for the solution of the problems under consideration, and many of these methods were graphical. During the ten years which have elapsed since then, the graphical methods have almost all been abandoned, as their inferiority has become evident, and at the present time the work of the Laboratory is almost exclusively arithmetical."

To compare with this opinion, let me quote some recent remarks of Millikan[1] which are in marked contrast to those of Whittaker and Robinson: "This value of the electron is also that at which Birge finally arrives as a result of his survey of the whole field of fundamental constants. It is true that he reanalyzes for himself my individual oil-drop readings and weights them so that he gets from them the value 4.768 ± 0.005 in place

---

1. Loc. cit.

of my value 4.770 ± 0.005, a result that is so much nearer mine than my experimental uncertainty that I am quite content - indeed gratified - but I may perhaps be pardoned for still preferring my own graphical weightings, since I thought at the time, and still think, that I got the best obtainable results in that way from my data. The person who makes the measurements certainly has a slight advantage in weighting over the person who does not, and the graphical method by which I got at my final estimated uncertainty is, I think, in the hands of the experimenter himself more dependable than least squares." In the same vein as that of Millikan, we have the words of the late Lord Rayleigh saying that the method of least squares is a good thing to read up on and then forget.

Degree of Assurance

Enough has been said to show that in every step of the formal process of discovering a law of relationship, the human element enters: first in the choice of the data, second in the choice of the assumed functional relationship, and third in the choice of the method of estimating the parameters. It is also significant to note that men of the calibre of Whittaker, Millikan, and the late Lord Rayleigh may choose differently. Scientific method in this sense fails to provide a universally accepted ultimate criterion to show that the result of one choice is better than another.

In the absence of any ultimate criterion, let us consider briefly the factors which must influence our assurance that a given theoretical relationship is approximately the corresponding objective relationship. In the acquisition of good data, it is obviously necessary for the experimentalist to be familiar with all relevant facts. For example, if the measurements involve the determination of a relationship between two physical properties of a material in a highly evacuated state, it is necessary that the experimentalist be familiar with the technique of high vacuum measurements.

Furthermore, a knowledge of available theory plays an important role. For example, one might investigate the relationship between pressure and volume of a gas at a given temperature and thereby come experimentally to a conclusion that the pressure varies inversely as the volume. However, the knowledge that this relationship is consistent with modern kinetic theory tends to increase one's assurance beyond that given by the observed relationship. In a similar way, the knowledge that Newton's laws of motion are, as it were, first approximations to more general laws of motion consistent with modern dynamical theory, tends to strengthen our assurance in the belief that Newton's laws are approximations to objective natural laws. In the last analysis, it is essential that the theoretical relationship be consistent with all known phenomena insofar as possible. In other words, the particular relationships under consideration should fit into available theory.

Let us pause to note the significance of these results. If good data depend so much upon the broad training of the investigator in respect to both empirical facts and theory, what reliance can we put in a large part of the great quantities of engineering data taken in a routine way by laboratory assistants, as so often is the case for example in the inspection of the quality of product from raw material to finished article?

B. Statistical Relationship

Let us again start with the case of two variables X and Y. If these are statistically related, it follows from what has been said that there is an objective probability p' of an observed pair of values falling within a given interval dXdY expressable in the following functional form:

$$p' = f'(X, Y, \lambda'_1, \lambda'_2, \ldots, \lambda'_i, \ldots, \lambda'_{s_1})dXdY. \qquad (8)$$

In the last analysis the law to be discovered involves the discovery of the function f' and the s_1 parameters in (8).

A little investigation shows, however, that any attempt to discover the relationship expressed in this particular way requires a comparatively large number of observations of pairs of values oftentimes running into the thousands. For example, the number of observed points in Fig. 6 showing the observed relationship between modulus of rupture and density of specimens of Sitka spruce is not sufficiently large to make it possible to discover the statistical relationship expressed in (8). What is usually done in such a case is to express the relationship either in the form of some coefficient such as the correlation co-efficient or in the form of a curve of regression of the form

$$Y = \varphi'(X)$$

giving the expected value of Y for a given value of X.

Good Data

The first essential characteristic of observed data used in the discovery of statistical relationship is that they should exhibit statistical control in the sense that the observed fraction p falling within the interval dXdY should approach p' (8) as a statistical limit. Here in the Laboratories we have chosen to call this condition that of statistical control. Important criteria for detecting whether or not a given set of data are controlled have been developed[1].

It may be shown that each of the two variables must satisfy the condition of statistical control. Hence through the application of these criteria, we may determine within closer limits than has heretofore been possible whether or not a given set of data are good in the sense that they may be used in the discovery of the statistical relationship. These contributions of the Laboratories constitute a definite important step in the discovery of statistical relationship by making it possible to get good data to begin with.

---

1. These are described in full in the book Economic Control of Quality of Manufactured Product, loc. cit., hereafter referred to as E.C.

Discovery of Relationship

There perhaps would be little question as to whether or not a statistical relationship existed for the data of Fig. 6. Let us therefore consider the data shown in the scatter diagram, Fig. 8. Is there a relationship in this case? The question is not easily answered at least by observation. It is beyond the scope of the present paper to go into the various methods of testing for the existence of relationship. We shall, however, go far enough to indicate the inherent difference between the test for a statistical and that for a functional relationship.

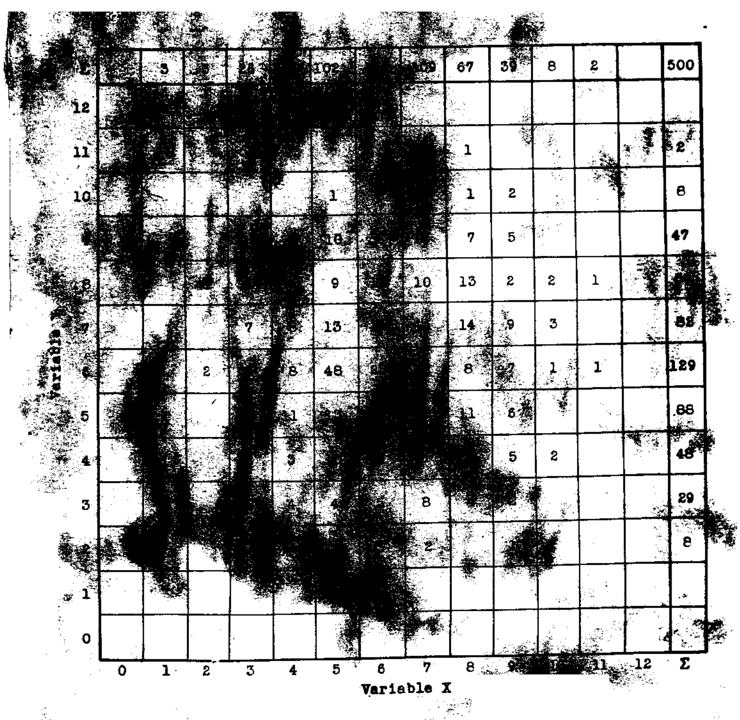

Fig. 8

Since observation of the data in Fig. 8 would lead us to believe that the regression is linear if there is a correlation, we shall consider here only the use of the correlation coefficient r as a measure of this relationship, it being recalled that this coefficient is given by the following expression:

$$r = \frac{\Sigma xy}{n\sigma_x \sigma_y} \tag{9}$$

where $x = X - \overline{X}$, $y = Y - \overline{Y}$, σ_x is the standard deviation of X, and σ_y is the standard deviation of Y. Of course, the objective value r' is zero if there is no correlation but linear regression. However, even though the objective value r' of correlation coefficient be zero, the observed value of r need not be zero. In fact it will be distributed about zero. In other words, the point to be considered is that the answer to a question as to whether or not the observed correlation r in a given case is significant must be considered in terms of the distribution function

$$f'(r', n) = 0 \tag{10}$$

of the observed correlation coefficient expressed in terms of the objective value r' of the correlation coefficient and the sample size n. In the particular case in hand, the observed correlation r is .0727. A knowledge of the distribution function (10) for the case r' = 0 shows that an observed value of r as large or larger than .0727 may be expected to occur approximately 11 times in every 100. In other words, we could expect to observe a value of r as large as .0727 a large fraction of the number of trials even though no correlation exists. What shall we conclude in such a case?

In the first place it should be noted that an appeal to the use of distribution theory enables us to test an hypothesis that the observed correlation coefficient came from a universe having a given value of r'. It does not, however, tell us anything specifically about whether the assumed universe is the objective one. Furthermore, our conclusion as to whether or not a given value of r is significant, depends upon what limits we set on the correlation coefficient. The answer to our question is therefore inherently indeterminate.

What we find in an attempt to discover whether or not two things are statistically related by means of the correlation coefficient is inherently the same as that which we find if we try to use any other available test - the best that we can do is to measure the likelihood of a given hypothesis but this is not a direct answer to our question.

Estimate of Parameters

Just as there are many ways of estimating the parameters in the functional, so also are there many ways of estimating those in the statistical relationship. Possibly we would be justified in saying that there is comparatively little basis for choosing between the different methods of estimation in the case of large samples, except that some are far more efficient than others, and for this reason should be used in order to reduce the cost of the investigation by minimizing the number of observations necessary to attain a given precision. For the case of small samples, as we usually have in engineering work, the problem of comparing the various methods becomes much more complicated. It may be shown[1] that under these conditions it is essential that we use certain methods for estimating the parameters, particularly in connection with the conditions which we meet in the measurement and inspection of the physical properties of raw materials and manufactured product. In other words, the formal part of the theory in the case of the statistical relationship limits the choice much more than it does in the corresponding case of the discovery of the functional relationship.

Degree of Assurance

It is customary to try to make use of probability theory in measuring the degree of assurance that we have attained in a statistical investigation. This theory is closely tied up with the theory of causes and is naturally linked with the name of Bayes through the theorem which bears his name. We cannot do more than indicate the nature of the limitations involved in trying to derive from probability theory a measure of the assurance of a given hypothesis.

- -

1. E.C., loc. cit.

A careful survey of probability theory shows that the best we can hope to do is to measure the likelihood that some function of a set of n observed values will have a value lying within a specified range upon the basis of the assumption that the observed sample has been drawn from some kind of a universe. This is true even if we make use of a posteriori probability. Our conclusion is limited, therefore, by the choice of hypothesis that we make. Hence, probability theory does not provide a basis for calculating the probability of a given hypothesis as compared with all others even in the case of statistical relationship. It appears that the best we can hope to do is to determine how likely a given set of data is in respect to a certain function of this set of data based upon one of an indefinitely large number of possible assumptions.

C. Empirical Relationship

As already noted, the empirical relationship may be either quasi-functional as that of the growth curve or the time series, or it may be quasi-statistical as that of the relationship between modulus of rupture and specific gravity of various kinds of woods shown in Fig. 7. In general, we have no definite, recognized, and accepted method of trying to discover relationships of this kind beyond that of finding some empirical curve or smoothed set of data in the case of the quasi-functional relationship. For the quasi-statistical relationship, there is no available method for making use of the theory of probability even in calculating the likelihood of the observed set of data based upon a given hypothesis. The methods of smoothing and fitting empirical curves in such instances do not contribute much to our assurance of a given form of relationship beyond that suggested by the observed data.

It is of interest to consider again the observed empirical relationship between modulus of rupture and specific gravity of different kinds of woods. In the publication already referred to, curves are given as shown in Fig. 9.

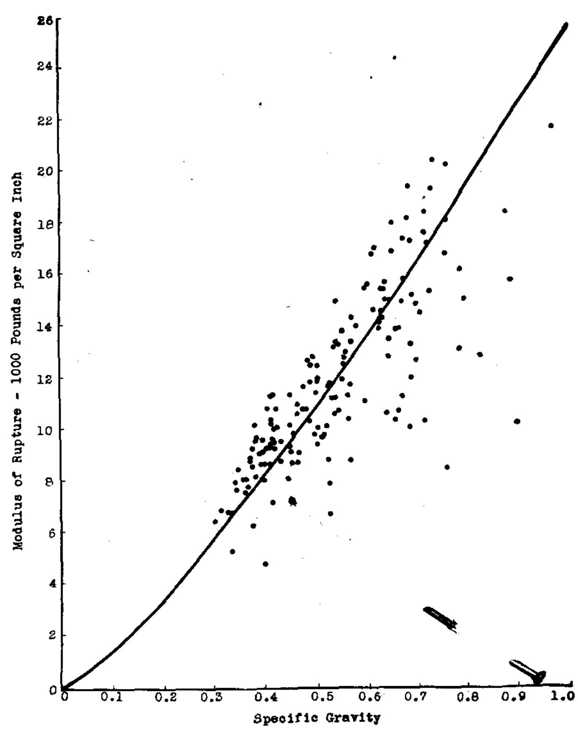

Fig. 9

What can be the significance of these curves? Certainly they have no significance in the sense of indicating functional relationship because it is apparent that no such relationship exists. In the second place, they cannot be curves of regression except in the case of repetitions as nearly as possible of tests on the same number of pieces from each of the total number of species[1].

SURVEY OF RÔLE OF JUDGMENT IN DISCOVERY

It will have been noted that the application of the scientific method in discovery involves a human choice at every step. For example, in the discovery of a functional or statistical relationship, the following choices must be made:

1. Choice of data.
2. Choice of functional form.
3. Choice of number of parameters at least in certain cases.
4. Choice of method of estimating parameters.

To a certain extent this field of choice is a kind of methodological No-Man's Land.

History of science shows, however, that the discoverers of the past have, in general, been those broadly trained in the particular field of discovery of their choice.

- -

1. True, a regression curve with any degree of nicety might be fitted to this set of points just as well as one could be fitted to the corresponding set of points, Fig. 6, representing the observed relationship between the same two physical factors for the case of Sitka spruce. In the first case, however, the curve would have no apparent physical significance whereas in the latter case it would have.

They have been those familiar with the status of experimental theoretical results in their particular field. The importance of theory in helping one to choose the right thing to be discovered is illustrated by the fact that several elements in the periodic table have been looked for and found because their existence was suggested by the blank spaces. So it is that many of the discoveries of science have been suggested by theory.

Furthermore, it is of interest to note in reading the history of science that important discoveries have usually come only after the investigator has surrounded himself for a considerable period of time with the facts bearing upon the subject and during this period has kept these more or less constantly in mind. It is true, however, history also indicates that many of these discoveries have only come after the investigator has dropped the search for a time more or less completely from his conscious consideration. In all cases, however, it appears that the preliminary conscious attention to the facts in hand is essential.

For fear that some one might falsely conclude that the formal parts of the scientific method do not add much assurance in any case, we need consider another example. In many branches of science, we often find cumulated frequency distributions plotted on some kind of probability paper. If the data so plotted lie on a straight line, it is concluded that the universe from which the data likely came is functionally of the form of that used as a basis for the grid of the paper.

Fig. 10 gives one such plot for a series of 15,050 measurements representing efficiencies on as many telephone instruments of a given kind. What conclusion would you draw? Do the points lie sufficiently close to the line to satisfy you that the objective frequency distribution is such as to give a straight line? If you answer yes, your eyes have misled you, for it can be shown by analytical methods that it is very unlikely that the objective universe is one of the same functional form as that used as a basis of the grid[1].

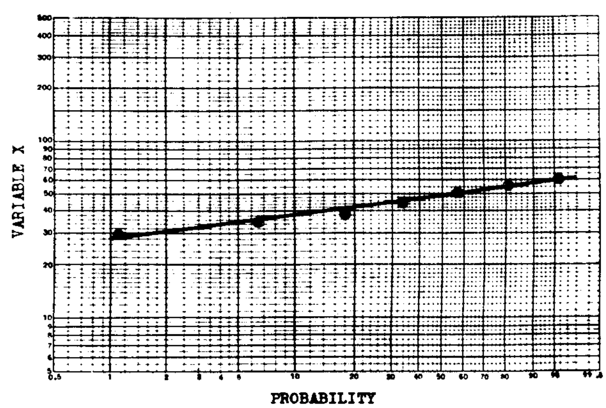

Fig. 10

- -

1. Goodrich, R. D., "Straight Line Plotting of Skew Frequency Data", Transactions of American Society of Civil Engineers, Vol. 91, p. 1, December, 1927.

LOGIC OF DISCOVERY IN RELATION TO PRODUCTION ENGINEERING

Thus far we have considered the problem of discovery as an object in itself with a view to indicating the relative importance of the human element and the formal steps of the scientific method in the process of discovery. Let us now consider briefly the problem of doing what we want to do time and time again as in the production of large numbers of the same kind of things.

We can only touch here upon some of the fundamental elements of this problem. In the first place it is essential that we adopt certain standards or specifications on the qualities of materials and finished products. In most instances it is true that the qualities of units of the same kind or the qualities of materials of the same kind are inherently statistical in nature in that no matter how hard we try to make things with the same quality or to choose pieces of the same kind of material all having the same quality, we may expect to meet with failure. Looked at, therefore, from the viewpoint of control of a given quality X, it is reasonable to believe that there are some limits within which variations in this quality should be left to chance provided we are not to waste time unnecessarily in looking for trouble when it is reasonable to believe that we cannot find it.

It is in this connection that the recent developments in the application of statistical theory give promise of being of great economical importance in that they enable an engineer to set limits on a quality or on any statistic θ of that quality, Fig. 11, such that the variations within such limits should be left to chance. The rational basis for establishing these limits is discussed in detail elsewhere[1]. The interesting thing is that they can be set so that over a long period of time they will insure economic control.

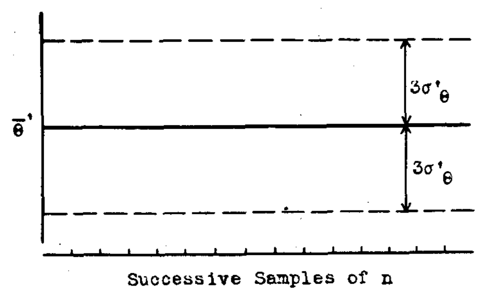

Fig. 11

In the more general case, the quality of the material for all practical purposes depends upon a set of let us say m physical and statistically related characteristics $X_1, X_2, .., X_i, .., X_m$. The standard of quality in such a case is inherently a statistical equation of state represented symbolically in the form

$$f'_s(X_1, X_2, .., X_i, .., X_m) = 0. \quad (4)$$

- -

1. E.C., loc. cit.

Assuming that it is essential to control each of the m characteristics, it is simply necessary to establish control limits for each of the m variables. It is also possible and often desirable to establish similar limits on the correlation coefficients between the pairs of the m different quality characteristics. In the case of each of the quality characteristics or correlation coefficients so treated, a deviation outside of the economic limits indicates the presence of causes of variation which should be eliminated.

It is interesting now to look again at the relationship between strength of wrought iron and temperature shown in Fig. 3. Suppose we were assigned the problem of producing wrought iron controlled in respect to the relationship indicated in this figure. This problem is analogous to many that we experience in the production of materials and apparatus of one kind or another. In the first place experienced men in the production field of wrought iron inform me that it is very unlikely that two pieces of the material would give exactly the same relationship as shown in the above figure. The question therefore arises as to what we should consider as a controlled product in respect to this particular quality.

In the light of what has been said, this problem reduces to an extension of the simple one of controlling a given quality characteristic about the expected value as illustrated schematically in Fig. 11. After we have done everything feasible to eliminate causes of variability other than those which should be left to chance, it is reasonable to expect that the values of strength corresponding to a given temperature would be distributed in some statistical distribution about some expected value different, in general, for each value of temperature. In fact, it is quite possible that the expected values of these frequency distribution functions would lie on some curve such as that shown in Fig. 3.

The suggested method of control in this case is to establish limits through the aid of Criterion I described above for each temperature in which we are interested. The general case is illustrated schematically in Fig. 12, showing only one of the several sets of limits.

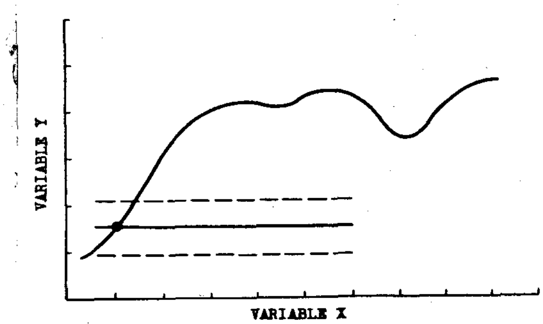

Fig. 12

Incidentally it is of interest to note that the method of presenting data by means of a smooth distribution or by a successive series of points indicating only the average values of one variable Y corresponding to another variable X as shown previously in connection with the properties of Bismuth alloys, is quite unsatisfactory from the viewpoint of control because it does not give any basis for establishing tentative limits of the type illustrated in Fig. 12.

CONCLUSIONS

Our brief survey of the field of scientific discovery has tended to emphasize the importance of the human element and of experience and training in discovery. It has indicated the way in which a knowledge of the formal processes involved in the scientific method may be of material assistance to the experimentalist in laying proper emphasis on each step of discovery in a particular case. One very important thing which cannot be stressed too much, because of its engineering significance, is that the customary emphasis on number of observations is in many instances entirely misleading.

Passing to the problems involved in the production of controlled quality in manufactured product, attention has been directed to the economic significance of recently developed criteria for indicating the presence of causes of variation which should not be left to chance. It is extremely significant that these criteria can be established in a way such that in the long run they serve to guide inspection effort at any stage in the production process from raw material to finished product in an economical way.