IMPORTANCE OF SOME STATISTICAL CHARACTERISTICS OF A STANDARD OF QUALITY

by

W. A. Shewhart
Bell Telephone Laboratories

Paper presented at Christmas meeting of the Econometric Society in Detroit, December, 1938.

Here I am, tonight, like the proverbial cat in the strange garret. Here I am as an engineer talking to you as economists and business statisticians about standards of quality. As an engineer, my experience is limited for the most part to those aspects of quality that you ignore. You talk about the demand for or the utility of a thing: I talk about those quality characteristics that make a thing what it is. You talk about the relation of a thing to one or more persons: I talk about the quality characteristics of a thing that are supposed to be independent of human beings. You presumably talk about subjective quality: I try to talk about objective quality.

However, when younger, I did frequent the writings of Doctor and Saint among economists and philosophers who talk about utility and the theory of value. After reading great argument, I did not, like Omar Khayyam, come out by the same door wherein I went. I had hoped to discover some quantitative way of relating wantableness, or what some of you call utility, to the physical characteristics of the thing wanted. I am still too young to be completely disillusioned and I feel that some of the things shortly to be said about the objective quality of a thing may contribute a wee mite to the understanding of that somewhat mystical attribute you call utility. But I would not think of taking your time today, busy as you are trying to find the royal road to freedom from our economic ills, if it were only to say something that might be of interest to you as students of utility of goods in the sense usually considered by economists, business statisticians and some philosophers. They, like the engineer of the past, have for the most part failed to take full account of the work of the little demons of chance that prevent man from making two things identically alike. I think there is something pertinent to be said about the comparative utility or wantability of different kinds of variability produced by these little demons of chance. For example, it appears that maximum utility of certain kinds of raw materials can only be attained by securing a statistically controlled variability.

That last sentence is rather wordy. Maybe it doesn't mean much to you. Let me therefore introduce without more generalities, a couple of performers on the economic stage that you, as well as engineers, are interested in. You talk and write about them and so do engineers. All of us feel that it is important to study their action in dealing with one another. Anything that affects this action is of interest to us. Hence I feel justified in talking to you briefly about the way in which a better knowledge of the habits of the demons of chance would likely change the actions of these two performers to whom I now direct your attention: they are the consumer and the producer.

Presumably it is the "want" of some consumer (or consumers) for a thing that gives rise to its utility. Still talking somewhat loosely, we may argue that the consumer's knowledge of the physical quality characteristics of the thing influences his want for the thing. For example, we find an economist saying: "The ignorance or indifference of the average consumer frequently leads him to purchase goods which he would not be willing to buy at all or at least would be willing to buy only at a much lower price <u>if he knew their real quality.</u>"*

Realizing the importance of having a basis for judging what is here referred to as real quality, corporate consumers and producers early in this century saw the need for establishing standards of quality in terms of measurable physical and chemical characteristics. The engineer and natural scientist took the lead in initiating the development of such standards. It was not long, however, before purchasing and sales agents began to make use of standards as a basis for contractual agreements and gradually the advantages of having standards of quality as a basis for corporate buying and selling have become recognized by the commercial as well as technical departments of corporate bodies. More recently there

- -

*S. H. Slichter, <u>Modern Economic Society</u>, Henry Holt and Co., p. 544, 1931. Underscoring mine.

has been a movement both here and abroad to extend the use of standards of quality to the case where either or both the consumer and producer are individuals. Some consumer groups have gone so far as to recommend the fixing of standards by legislation.

All such effort implies that standard quality can be specified in such a way that one can definitely determine whether or not a given thing is of standard quality. However, within comparatively recent years, engineers and scientists have begun to see the limits imposed by the demons of chance on the process of making and using standards of quality.

As yet economists and business statisticians (as well as other social scientists) have not, so far as I am aware, given much consideration to the importance of these limitations. For example, many still plan purchasing and selling programs as though a standard could be defined in an operationally verifiable manner. This situation is to be expected until engineers and natural scientists provide economists and other social scientists with an accurate description of what constitutes a standard of physical quality and of the essential factors in the process of judging whether or not a thing is of standard quality.

Moreover the economist must look to the natural scientist to tell him what he may expect in the way of variability in quality at the hands of the demons of chance under different conditions of technical control. He needs this if he is to help the natural scientist in determining the wantability or demand for different degrees of variability. Of course, I assume that the technique of attaining such control, as such, is no part of the subject matter of economics, but knowledge of the existence of a technique of various potentialities helps to condition the possible demands of consumers for any given product and hence influences the adaptations of human effort that it is the business of economics to study. The business statistician and economist must also know something about what kind of human effort and what kind of production system is required in the control of quality to attain a specified degree of control of the variation if he is to be in a position to study intelligently the relationship between different degrees of control in production and the technical and social environment. For example, if the maximum degree of control is desired, the production unit cannot be made too small and there must be close cooperation between the three major steps - specification, manufacture, and inspection in the production of goods.

There are, in fact, the following four questions to which the economist and business statistician as well as some other social scientists must turn to the engineer for answers:

1. What are some of the limitations to the specification of a standard of quality?
2. What degrees of control of variability can be obtained?
3. How can control of variability be attained?

4. What are some of the limitations to judging whether or not a given thing is of standard quality?

In what follows, I shall very briefly consider these four questions from the viewpoint of the effects of chance or unknown causes of variability.

LIMITATIONS TO SPECIFICATION OF STANDARD QUALITY

Let us see how far we can go in specifying standard quality in the following form:

A The quality of a thing in order to be considered as standard shall be such and such.

From the viewpoint of logic, it should be noted that such a specification states a condition that the quality of a thing must have *if* it is to be called standard. It does not affirm that a thing having this quality exists or that it can be produced, although in practice needless to say, an attempt is made to write specifications that can be met. Now, if such a specification of quality is to be definite, it must be operationally verifiable. Let us see what limitations are encountered in attaining such definiteness.

The Meaning of Quality

We must start with a word or two about the meaning of the quality of an object such as any piece of apparatus or raw material. As already noted, a thing, if wanted by some person or group of persons, is said to possess wantableness or utility. This utility depends upon the relation between a human being and the physical thing. In addition to this subjective quality or utility, there is the objective quality of the thing expressible in terms of its physical and chemical properties. For our present purpose, these are assumed to be independent of human interest or volition.

I shall assume that any quality characteristic has operationally verifiable meaning only in the sense of one or more operations of experiencing or measuring the quality of a thing. A simple illustration will help to make this point clear. Let us consider what is perhaps one of the simplest quality characteristics, namely, length. For example, the length in question might be the distance at 0° C. between two parallel lines engraved upon the platinum-iridium bar deposited in the International Bureau of Weights and Measures and declared to be the international prototype of the meter by the First General Conference on Weights and Measures, this bar being subjected to normal atmospheric pressure and supported by two rollers not less than a centimeter in diameter situated symmetrically in the same horizontal plane and at a distance of 571 mm. one from the other. The operationally verifiable meaning of this length involves one or more operations of measurement. Thus we might measure it in terms of the British Standard Inch. Reliable measurements of this kind were carried out with the utmost care in 1895, 1928, and

1933 and gave the following results: 39.370113, 39.370147 and 39.370138.

Enough has been said to illustrate three aspects of any quality characteristic: 1) the concept, in this case the concept of length; 2) some operation or operations of measurement; and 3) a number obtained in the operation of measurement.

For our present purpose, it is important to classify operations of measurement as being either nondestructive or destructive of the object measured. Examples of nondestructive measurements are those of length and mass of a steel rod, the capacity of a condenser, and the electrical resistance of a wire. Examples of destructive measurements are tests for the blowing time of fuses, tensile strength of materials, and chemical composition of foods, drugs, and the like, to mention only a few. Let us now briefly examine these two kinds.

Let us start with the class of nondestructive measurements illustrated by the measurement of length in inches of the physical standard of the meter. We often conceive of this length as being constant and having some one fixed value under the physical conditions specified. From the viewpoint of verification by measurement, however, the length does not have one unique value. In fact any operation of measurement may be repeated again and again at will and such a sequence of repetitions will generate an infinite sequence of observed values,

$$X_1,\ X_2,\ \ldots,\ X_i,\ \ldots,\ X_n,\ X_{n+1},\ \ldots,\ X_{n+j},\ \ldots \qquad (1)$$

Irrespective of the care used in making such a sequence of measurements even under conditions of the most careful laboratory control, the observed values will not, in general, be the same. This fact is illustrated above by the three measurements of the standard meter in terms of the inch. Hence an objective quality characteristic, such as the length of an object, expressed in terms of a given operation of measurement, is not a single value but instead is an infinite sequence. In fact, the verifiable constancy of any quality in terms of measurement exists only in the sense of the reproducibility exhibited by such a sequence.

To understand the sense in which a quality characteristic has measurable objectivity, we must also take into account the fact that there are customarily assumed to be an indefinitely large number of methods or operations of measurement. This is obviously true for the simple characteristic length. Thus the meaning of a quality characteristic in terms of measurement must find expression not simply in terms of one infinite sequence corresponding to one operation of measurement but instead in terms of an indefinitely large number of sequences corresponding to the indefinitely large number of methods of measurement of a given quality assumed to exist. We shall refer to this as the operational meaning of a quality characteristic X and shall represent it schematically as follows:

$$X \xrightarrow{O} \begin{cases} X_{11}, X_{12}, \ldots, X_{1i}, \ldots, X_{1n}, X_{1n+1}, \ldots, X_{1n+j}, \ldots \\ X_{21}, X_{22}, \ldots, X_{2i}, \ldots, X_{2n}, X_{2n+1}, \ldots, X_{2n+j}, \ldots \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ X_{k1}, X_{k2}, \ldots, X_{ki}, \ldots, X_{kn}, X_{kn+1}, \ldots, X_{kn+j}, \ldots \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \end{cases} \quad (2)$$

Next, let us see in what sense a specified quality characteristic that can be measured only by a destructive operation possesses verifiability. Here again there may be an indefinitely large number of operations of measurement. The application of such a test to a given object, however, leads only to a single value instead of the infinite sets of values shown in (2). This fact is of importance from the viewpoint of specification, as we shall soon see.

Specification of Standard Quality

Let us start again with quality characteristics measurable by nondestructive operations. A classic example is the specification of linear dimensions. About 1870 the engineer introduced the concept of tolerance limits and since then he has as a rule specified for a given quality characteristic X that it shall lie within certain stated tolerance limits L_1, L_2. It no doubt seemed to engineers then as it does to many engineers as well as others today that such tolerance limits are the "cats whiskers" for specifying quality in an operationally verifiable way. Let us therefore see in what sense a specification that a quality characteristic X shall lie within specified limits L_1, L_2 in order to be standard is subject to verification.

Keeping in mind that we are still considering non-destructive operations of measurement, the operational meaning of X is given by the set of infinite sequences in (2). No mental effort is needed to see that the specification as stated is indefinite until we are told what is to be taken as the one observed value of X that is to lie between L_1, L_2. But you may say that such a specification implies that all or a given fraction p' of all the values of X in (2) must lie within L_1, L_2. If so, however, there would be no way of practically verifying such a meaning because we could never observe all the values of X even though we admit that such meaning is theoretically verifiable. Hence a specification of a quality characteristic in terms of tolerances is not operationally verifiable in a practical sense if it is to include the whole meaning as defined in (2).

To make such a specification definite it is necessary to specify the operations to be used in measuring X, the number of observations to be made by each method, and the function ψ of these observed values that is to lie within the limits L_1, L_2. In practice one operation of measurement is often specified and the arithmetic average of the n observed values of X corresponding to n repetitions of this operation of measurement is required to lie within L_1, L_2. More generally, however, we might specify k operations of measurement and n_1, n_2, ..., n_i, ..., n_m repetitions respectively for these methods. For all n's equal, the required observations might be indicated schematically as in (3) by means of the rectangle.

$$X \xrightarrow{0} \left(\begin{array}{|l|l} \cline{1-1} X_{11}, X_{12}, \ldots, X_{1i}, \ldots, X_{1n}, & X_{1n+1}, \ldots, X_{1n+j}, \ldots \\ X_{21}, X_{22}, \ldots, X_{2i}, \ldots, X_{2n}, & X_{2n+1}, \ldots, X_{2n+j}, \ldots \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots & \cdots\cdots\cdots\cdots\cdots\cdots \\ X_{k1}, X_{k2}, \ldots, X_{k2}, \ldots, X_{kn}, & X_{kn+1}, \ldots, X_{kn+j}, \ldots \\ \cline{1-1} \multicolumn{2}{l}{\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots} \end{array} \right. \qquad (3)$$

If by trial, the arithmetic mean of these is found to lie within L_1, L_2, then the requirement of the specification would have been met. Hence such a specification is operationally verifiable in a practical sense.

The important point to be noted is that such a specification does not cover the whole of the operational meaning of X. Furthermore, the part of the operational meaning represented by the portions of the sequences outside of the rectangle in (3) is usually just as important or perhaps more important than that within. We shall see how significant this fact is later when we consider the problem of judging whether or not quality is standard.

It should also be noted that instead of specifying tolerance limits on one function ψ, such as the arithmetic mean, of the observed values within the rectangle (3), one might specify s such functions and their tolerance limits

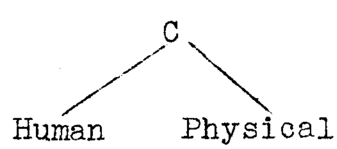
$$\psi_{1i} \leq \psi_i \leq \psi_{i2} \qquad i = (1, 2, 3, \ldots, s) \qquad (4)$$

within which the values of ψ_i must lie in a given case in order to be considered as standard.

Thus far we have considered the limitations to the specification of the operational meaning of a quality characteristic only in terms of the observable values. Associated with each one of these potentially observable values, there is a condition C of measurement in the sense that the word "condition" is used in the phrase "same essential conditions." Thus instead of a simple infinite sequence (1) corresponding to one operation of measurement we really have

$$\begin{array}{ccccccccccc} X_1, & X_2, & \ldots, & X_i, & \ldots, & X_n, & X_{n+1}, & \ldots, & X_{n+j}, & \ldots \\ | & | & & | & & | & | & & | & \\ C_1 & C_2 & & C_i & & C_n & C_{n+1} & & C_{n+j} & \end{array} \qquad (5)$$

It is obvious that any condition C_i involves a human and a physical element that we may symbolize as indicated below:



Customarily we impose the requirement that the conditions of measurement shall be essentially the same but this apparently involves a human subjective judgment, and hence cannot be made operationally definite. Perhaps the nearest approach to definiteness would be to specify the person or persons that are to exercise the judgment that the conditions of measurement are maintained essentially the same.

Now a word is in order about specification of quality in terms of operations of measurement that are destructive of the thing measured. The first thing that has to be done is, as before, to choose the operation of measurement. Let us assume that this is the k th. Then the specification in terms of tolerance limits takes the simple form,

$$L_1 \leq X_{k1} \leq L_2 \quad , \qquad (4.1)$$

where X_{k1} is the one and only measurement to be taken on the quality of the thing in question and (4.1) implies that this value X_{k1} if and when taken must fall within the tolerance limits L_1, L_2 in order for the quality of the thing to be standard.

<u>Specification in Relation to Consumer and Producer</u>

Again let us start with nondestructive operations of measurement. In general, the consumer is interested in the quality of a thing throughout its life. He may choose to specify the quality in a way that can be verified in practice by taking a block of measured values such as shown enclosed within the rectangle (3) and by checking to see if the values of the specified functions ψ_i fall within the stated tolerance limits (4). Even if he does, however, this does not mean that he is not interested in the quality theoretically verifiable in terms of measurements outside this rectangle. In fact, he is interested in the results of any limited series of tests only insofar as they reveal something about the quality operationally expressible as in (2). The statistician would perhaps describe this situation by saying that the results of taking any finite set of measured values such as those shown within the rectangle of (3) constitutes only a sample of those that may be taken and that are of interest to the consumer. The consumer can draw only a probable inference from such a sample about what he may expect to get if he were to carry out further measurements.

Let us compare this situation with that existing for destructive operations of measurement. It is obvious that the whole of the product going to a consumer could not be examined to check whether or not its quality X falls within specified tolerance limits such as (4.1) without destroying the whole of the product. Nevertheless the consumer wants to have assurance that this quality, if tested, would lie within the specified limits. To provide a basis for judging whether or not the thing as a whole would, if tested, meet the specification, the consumer often specifies that a sample shall be taken and tested and that the result thus obtained shall fall within tolerance limits L_1, L_2 (4.1). However, even though the quality of the sample satisfies this requirement, one can only draw a probable inference about whether or not the untested portion of the thing would meet the same tolerance.

Now we are in a position to spot a close similarity between specifications in terms of nondestructive and those in terms of destructive operations of measurement. In neither case is it feasible to write a practically verifiable specification of the quality of a thing such that even though satisfied by the quality of a thing we may be sure that the quality of that thing or part thereof as received by the consumer is what is wanted (even assuming that he knows what he wants). This follows from the fact that we cannot write such a specification in terms of nondestructive operations of measurement because in order to be practical we must confine our attention to a finite set of data such as enclosed in the rectangle of (3); and we cannot write such a specification in terms of destructive operations of measurement because we cannot verify the quality of a thing without its destruction. Hence for both kinds of measurement, it is not possible to specify the quality of a thing wanted in a way that may be verified in practice. That is to say, one cannot specify the quality of a thing wanted in the sense of defining operationally the quality of the thing after it is in the hands of the consumer in such a way that one can determine with certainty that a thing possesses this quality at the time and after it reaches the consumer.

<u>Specification in Terms of a Sample and in Terms of the Whole</u>

We are now in a position to differentiate between any attempt to specify a standard for a quality characteristic in its complete operational meaning as indicated by (2) and any attempt to specify it in terms of a sample of these potentially infinite number of measurements. Likewise for specifications in terms of destructive operations of measurement, we note the possibility of specification in terms of either or both the sample and the whole.

I think it will be generally admitted that wantability on the part of a consumer is conditioned by the quality as a whole rather than as a sample. Hence a specification should attempt to define the quality wanted as a whole. Yet the consumer realizes that by placing certain tolerance requirements (4) or (4.1) on the sample he can increase his assurance that the quality of the whole would, if tested, meet similar tolerances. Hence the consumer often feels it necessary to include operationally verifiable requirements in terms of sample tolerances. He may in fact issue these in the form of inspection instructions defining acceptance rejection requirements.

From the viewpoint of economics, however, we should note that the use of the sample provides only a certain degree of assurance that the quality of product after it reaches the consumer is standard in the sense defined for the whole. Insofar as the same degree of assurance can be obtained in a cheaper way through the use of certain control techniques, the use of these techniques becomes of interest to economists studying the distributions of scarce means for satisfying given human wants. For example, it seems to be of particular importance to the economist and business statistician called upon for advice in helping to shape sales and purchasing programs to keep in mind that a specification defining what a consumer wants does not in itself provide a basis for assuming that the product accepted by the consumer will prove to have the quality defined in the specification. Insofar as product does not prove to meet the specification, the actual cost to the consumer must include costs of replacement. In studying the exchange of goods, it is therefore necessary to take into account not only the specification of what is wanted but also the assurance that the product exchanged will be found to have the quality specified.

What Fundamental Change in Content of Specification is Suggested by Statistical Theory?

We are now in position to consider two fundamental modifications in specifications suggested by statistical theory. One of these is to take account of the importance of the order in the sequences (2). For example, the statistician immediately thinks of the random order as having many desirable characteristics, particularly from the viewpoint of being able to make valid predictions upon the basis of a sample.

The other important modification suggested by statistical theory has to do with the requirements placed in a specification that must be satisfied in order for product to be acceptable to the consumer - that serve, in other words, to provide assurance that quality of product accepted will prove to be that specified as wanted. It has been shown elsewhere, for example, that these requirements should be made to depend upon the <u>kind of variability</u> existing in sequences such as (2), or, in other words, upon the degree of control of the demons of chance.

You may, of course, be saying that in the case of destructive operations of measurement we do not have sequences such as (2). True enough, we do not if we confine our attention to the measurement of the quality of a particular thing. Such measurement gives us only one value if the operation is destructive. There is, however, nothing to prevent us from measuring, as it were, the quality of a production process. Assuming that the process can turn out an indefinitely large number of things

$$O_1, O_2, O_3, \ldots, O_i, \ldots, O_n, O_{n+1}, \ldots, O_{n+j}, \ldots \qquad (5)$$

of the same kind, we could of course test each one of these by some specified method and in this way we would get an infinite sequence such as (1). Likewise we may conceive of the process as turning out not one such sequence but an indefinitely large number of such sequences,

$$
\begin{array}{l}
O_{11}, O_{12}, \ldots, O_{1i}, \ldots, O_{1n}, O_{1n+1}, \ldots, O_{1n+j}, \ldots \\
O_{21}, O_{22}, \ldots, O_{2i}, \ldots, O_{2n}, O_{2n+1}, \ldots, O_{2n+j}, \ldots \\
\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
O_{k1}, O_{k2}, \ldots, O_{ki}, \ldots, O_{kn}, O_{kn+1}, \ldots, O_{kn+j}, \ldots \\
\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots
\end{array}
\qquad (6)
$$

Now we might test all in the first row by one operation of measurement, all in the next row by another operation and so test each row by a different operation. The test of each object, being destructive, would yield only one number. Hence, as a result of carrying out such tests, we would get a set of sequences the same as shown schematically in (2).

It should also be noted that even when the operation of measurement is nondestructive, we may for some reason choose to take but one measurement for each object. If, however, we consider this one measurement not as a measure of the quality of the thing but as a measure of the quality of the operation of producing the thing we can set up a sequential order such as (6) among the objects produced. Then the set of all single observed values per thing could be arranged in the same sequential order as (6) and (2).

Now what significance from the viewpoint of specification has this possible shift from considering an observed value of quality as the measurement of the quality of a thing to the measurement of the quality of the process or operation of producing this thing? Insofar as the specification is to define what kind of thing is wanted by the consumer, no importance attaches to this shift. However, the possibility of this shift is of great importance from the viewpoint of stating requirements that must be met by a sample of a given lot of product in order to provide assurance that the quality of the product accepted by the consumer will prove to meet the specification of what is wanted. I think you will find that the prevalent practice is for the consumer to try to write specifications as though each lot of things submitted to him for purchase constituted what the statistician calls a "unique" sample. He may not include any requirements tending to fix the "origin" of the sample. He does not as it were ask for a "pedigree". In much the same way that it profiteth a breeder of fine animals to look not only at the animal but also at its pedigree so also does it often profit the consumer to look not only at the thing offered as satisfying his specified wants but also at the "pedigree" of the producer's production process for this kind of thing. We shall return to consider some of these advantages later.

WHAT KIND OF VARIABILITY IS WANTED?

Both the consumer and the producer are faced with the fact that they must live with variability of one kind or another. There are, however, an indefinitely large number of kinds of variability differing all the way from that exhibited by the throw of the die to that exhibited by the stock market. Before deciding upon the details of a specification, the consumer must decide what degree of control of variability he wants. Having a specification of what the consumer wants, the producer must decide how he can most efficiently control his production processes in order to attain the degree of control of the variability of the quality desired by the consumer.

"Exact" Control

Way back (1787) at the time of the introduction of interchangeable manufacture as the basis for mass production, corporate consumers and producers were imbued with the exactness of science and hit upon the idea of specifying with exactness the qualities of pieceparts. They conceived of making all pieceparts of a given kind so as to have the same dimensions so that these could be fitted together under the conditions of mass assembly. The only difficulty with this scheme was that no one succeeded at that time nor since in making things identically alike.

Control Within Tolerance Limits

By about 1870, engineers hit upon the idea of trying to make things so that for each specified quality characteristic X all things of that kind should lie within previously specified limits L_1, L_2 technically known as tolerance limits. For example, in specifying the standard of quality for a specific characteristic like length, two tolerance limits would be given and a thing would be considered to have standard quality if its length fell within these limits. In writing such a specification, it was more or less tacitly assumed that there was a unique observed value of length. Instead, as we have already noted, there are an indefinitely large number of infinite sequences of values represented by (2). Hence in order to make such a concept of specification operationally verifiable, it is necessary to define the segment of this sequence that is to be used as a basis for computing the functions ψ_i that are to lie within tolerance limits.

Control Within Minimum Tolerance Limits

As was to be expected, however, both producers and consumers soon became interested in the economic advantages to be attained by minimizing the tolerance limits. That is to say, they wanted to attain a condition where it was not only possible to hold the variability within tolerance limits but also to reduce the tolerance range as much as possible.

Requirements on variability expressed in terms of tolerance limits do not attempt to fix the variability within a sequence. Let us refer to the set of sequences (3) in which the portion of these sequences used as a basis for computing the function ψ_i on which tolerances are supposed to be set is enclosed within a rectangle. Early concepts of setting these tolerance limits paid no attention to the order in which the numbers occurred in any one sequence. Even when applied in the theoretical sense to the sequence as a whole, attention was given only to the limiting values in the sequence and not to the <u>order</u>. This resembles very much the early stress in statistical theory upon the observed frequency distribution of a set of numbers to the neglect of the order in which the numbers actually occurred.

<u>Control Giving Maximum Validity to Predictions</u>

It was early realized that insofar as specifications specify the quality wanted by the consumer, they cannot be operationally verified prior to the acceptance of the product. Even if the specification required that a sample be taken and imposed an acceptance rejection tolerance on the quality exhibited by such a sample, one still could not be sure that the accepted product would prove to have the quality specified as wanted. Consumers and producers have within recent years become aware of the fact that the degree of assurance provided by such sampling results depends upon the nature of the variability exhibited by the potentially infinite sequence. Perhaps one of the most important contributions of statistical theory has been to call attention to the vastly greater importance of the order in which observed values occur than of the frequency distribution of these numbers, particularly from the viewpoint of making possible valid predictions as to the nature of the fluctuation in an unexamined portion of an infinite sequence based upon an examination of a part of such a sequence. The statistician would be inclined to suggest, for example, that maximum validity of prediction can be attained by eliminating the causes of variability until resultant variability exhibits random order. It follows, however, that if the producer and consumer are to make use of this concept of randomness as a basis for specifying the quality wanted, it is necessary for the statistician to provide operationally verifiable criteria of randomness to be written into the specification.

It is needless to point out to this audience that we do not know of any unique way of specifying randomness in an operationally verifiable manner and so there is no unique way of writing such a specification. However, abundant evidence is available today to show that the practical techniques of statistical control of quality of product enable one to approach this goal in a perfectly definite manner.

Maximum Control

Another way in which engineers, producers, and consumers approach the subject of control of variability is to demand that everything possible be done to eliminate causes of variability. Such demands are particularly important in the control of the quality of goods wherein the failure to meet the specified requirements might lead to bodily injury. I have shown in considerable detail elsewhere* that it appears feasible to apply to good advantage what I have termed the operation of statistical control in the attainment of this goal. I have shown that it appears that through the use of criteria of control it is possible to detect causes of variability which are often overlooked when we simply try to repeat operations under essentially the same conditions.

Economic Control Within Tolerance Limits

It was stated earlier that both the consumer and producer are interested in knowing as much as possible about the possible kinds of control of chance causes of variability. Thus, even though the consumer might be interested simply in control within limits, the producer may reasonably be expected to ask himself how he can most efficiently attain this kind of control. This leads naturally to a consideration of what I

---

*Lectures given before the Graduate School of the Department of Agriculture, Washington, D.C., March, 1938.

have termed elsewhere* the economic control of quality. In principle it amounts to defining certain operations of applying, in addition to the tolerance limits, two control limits and an aimed-at value. Since the use of such a technique of control appears to provide control within tolerance limits at a minimum cost, it also becomes of interest to the consumer to see that such techniques are used in the production of his goods. He may, in fact, choose to require some evidence that such techniques are being used.

Maximum Control so as to Give $dy = f(X)\ dX$

In many cases the consumer may wish to attain a product the quality of which is not only statistically controlled but also controlled in a way to provide a specified functional distribution $dy = f(X)\ dX$, where dy represents the probability of the occurrence of a piece of product with the quality X lying within the interval $X - 1/2\ dX$ to $X + 1/2\ dX$.

Thus briefly we have surveyed some of the kinds of control of variability of quality that may be wanted by consumers and hence have surveyed some kinds that must be reduced insofar as possible to specifiable form.

---

*Economic Control of Quality of Manufactured Product, D. Van Nostrand and Company, New York, 1931.

HOW ATTAIN THE DESIRED DEGREE OF CONTROL?

As economists and business statisticians, you are perhaps no more interested in the technical details, as such, of the methods available for attaining different degrees of control than you are in the technical methods of producing highly complicated apparatus. However, you are interested in the study of the exchange of goods between producer and consumer. Assuming that consumers want a particular degree of control of variability of the quality of product it is presumably a matter of economic interest to determine what kind of a production system is the most effective means of satisfying such wants.

Let us assume, for example, that the consumer wants what has been called the maximum degree of control. Would a system of production in which there are a very large number of producing units, each of which has a small output, be as efficient an economic system as one in which there were a small enough number of units so that each could apply the techniques of mass production? To get at an answer to this question, we must look at the steps involved in making things to specification.

Three Fundamental Steps

Perhaps the simplest picture of the production of goods to satisfy human wants may best be thought of as involving the following three steps:

1. The specification of the quality of the thing wanted.
2. The production of things supposed to have that quality.
3. The inspection of the quality of the things produced to see if they have that quality.

We might, as is often done, approach the study of the exchange of goods as though the three steps were independent one of the other. This from certain angles is somewhat the ideal to be sought and theoretically it could be attained _if_ it were possible to take these three steps independently. The difficulty is that the steps cannot be taken independently if we are to be practical.

Of course, one could choose to specify the quality of the thing that he wants independent of whether or not it could be attained in the light of our present knowledge but such action is obviously not very sensible. For example, no consumer would likely think of specifying a perpetual motion machine even though he would like to have one for the simple reason that no one knows how to build one. In much the same way, it would be rather foolish to specify that one wanted a quality of product that was statistically homogeneous when considered in the order of its production if no one knew of a way of producing a product of such quality (at least at cost that would be within the means of the consumer). Obviously, therefore, the first step, specification, cannot be taken in a practical way without a knowledge of what is attainable in the second step.

In turn, however, we do not know what is attainable in the second step until a number of pieces of the desired kind of product have been turned out and inspected. Let us assume now that a number n of pieces of the given product has been made and that you have the job of determining whether or not some measurable characteristic X, supposed to be homogeneous when taken in the order in which the pieces of product were turned out, meets this requirement. Obviously, no statistician would place much reliance on any accepted test for homogeneity if the sample size n is small. This means simply that a considerable amount of product must be turned out before adequate evidence is available to indicate whether or not the causes of variability are in a state of statistical control. Experience shows, however, that it is almost a universal rule to find that such initial evidence indicates lack of homogeneity in the quality of any product in the initial stages of production. This means that it is necessary to go back into the process of production to detect and eliminate, if possible, assignable causes of variability, before it is possible to attain the desired state of control.

I trust that enough has been said to indicate in a nontechnical way the dependence of the three steps in the production of goods to satisfy human wants. So long as one limits himself to a consideration of ways and means of satisfying human wants that involve the control of variability of quality on the assumption that these three steps are independent, just so long may he expect to find that his conclusions do not fit the facts. If one is to be realistic, it is necessary to take into account the interdependence of these three steps and to view them pretty much as if they go around in a circle. In order to arrive at a basis for fixing an operationally definite specification of quality that can be met at reasonable cost with known methods of production, it is necessary to go around and around in this circle until we have attained adequate evidence that the little demons of chance that produce erratic causes of variability have been detected and eliminated.

It is significant from the viewpoint of determining the most desirable sizes of production units to note that experience indicates that one must go around and around this circle many times before he can attain a statistical state of control. In other words, such a state appears to be attainable only under the conditions of mass production where the operational techniques of statistical control are used in detecting assignable causes of variability and where the three steps, specification, production, and inspection, are closely coordinated.

Economists, of course, have again and again pointed out the advantages of mass production as a means of satisfying human wants efficiently. They have pointed out that under such conditions it is possible to apply the latest developments in science through the introduction and use in industrial research laboratories by large producing units. All such discussion, however, has been limited pretty much by tacit assumption that the application of science makes it possible to carry out the three steps in the production of goods to satisfy human wants in a more or less independent manner. Now, in what has been said above, we get a glimpse of another advantage of the process of mass production as a means of satisfying human wants. We see that if we are to go as far as it is practical to go in controlling variability in quality, it is necessary to employ mass production techniques and to subject these to what I have termed elsewhere the operation of statistical control.*

LIMITATIONS INVOLVED IN JUDGING QUALITY

Let us assume that it is feasible to specify the quality that a thing must have in order to be standard. How would you set about to determine if a given thing meets the specified standard? It is necessary that you make the following judgment:

B. This thing is (or is not) of standard quality.

This sentence is a proposition. It declares in effect that at

---

*See, for example, Economic Control of Quality of Manufactured Product, D. Van Nostrand and Company, 1931, and Statistical Method from the Viewpoint of Quality Control, United States Department of Agriculture, 1939.

least one thing exists that is of standard quality. Logically, this judgment B is fundamentally different from the specification A which attempts to define a standard of quality irrespective of whether or not any thing having this standard quality exists or can be made. For example, the proposition B, but not the specification A, has the property of being either true or false.

Some Aspects of the Problem of Judging

Let us think of some thing supposed to have a quality characteristic that remains "constant" as for example, the thickness of a gauge, the length of a measuring tape, or the capacity of a condenser, to name only a few. Let us assume that the measurement of one such characteristic X by some chosen method of measurement can be repeated again and again at will. This means that we may conceive of taking an infinite sequence of measurements. Any moment such as the present divides this potentially infinite sequence into two parts: the n measurements of X that have been taken and the rest of the measurements that (theoretically at least) can be taken. Schematically, the situation is that shown below (6).

$$X_1,\ X_2,\ \ldots,\ X_i,\ \ldots,\ X_n, \;\Big|\; X_{n+1},\ \ldots,\ X_{n+j},\ \ldots \qquad (6)$$

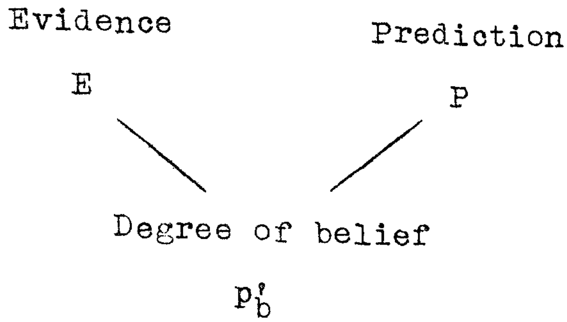
Past | Present | Future

For the more general case where we consider length as operationally defined in terms of all possible operations of measuring length (2), we could divide each infinite sequence as in (6).

Now, we may think of the present as marking the time at which the thing passes from the producer to the consumer. From the viewpoint of the consumer, the quality of interest is that to the right of the vertical line called the present in (6). Measurements appearing to the left of this are of interest to the consumer only insofar as they constitute a sample of what may be expected. Such data constitute evidence E upon which to base a prediction P that quality of the thing will prove to be that specified. Such an inference or prediction cannot be made with certainty: it is only probable upon the basis of the evidence E. I shall assume that there is an inherent degree of belief p_b' that it is rational to hold in the prediction P based upon the evidence E. Evidence E, prediction P, and degree of rational belief p_b' constitute three aspects of the knowledge that one has about the quality of the thing. These three aspects of knowledge of quality have been considered in some detail elsewhere.* Schematically the three aspects of empirical knowledge may be represented as follows:

---

*W. A. Shewhart, Statistical Method from the Viewpoint of Quality Control, U.S. Dept. of Agriculture 1939.



A point to be emphasized here is the difference between: a) verifying whether or not a thing has the quality specified and represented by the symbol P of prediction, and b) verifying a judgment which in this case is a prediction based upon evidence E.

Now, let us examine a little more carefully the significance of this process of judging from the viewpoint of the relation between consumer and producer.

<u>Importance of the Process of Judging from the Viewpoint of Producer and Consumer</u>

Let us assume for the sake of argument that a consumer is armed with a specification of the quality that he wants and that the specification is subject to operational verification at least in the theoretical sense. Let us also assume that the specification covers the quality of the thing wanted <u>after</u> it reaches the consumer. Now let us assume that several producers offer goods presumed to have the quality specified by the consumer. At the time of taking over the goods from any one of these producers the consumer can never be sure that the goods will prove to have the quality specified. The most that the consumer can do is to accumulate a finite amount of information to be used as a basis for judging whether or not it is likely that the goods accepted will prove to be what is wanted. Hence we arrive at the important conclusion that in purchasing goods a consumer must choose not between goods of qualities known with certainty but rather between goods with qualities known with different degrees of probability.

The consumer may, of course, decide to take the word of each producer as to whether or not the goods offered are of the quality specified or he may require inspection. If the consumer is interested, as he generally is, in minimizing the cost of the goods he will want to minimize the number of times that he accepts goods and later finds that the quality was not what it was expected to be. In other words, the consumer wants to use the technique that will help him to render valid judgments most often. This means that he is vitally interested in minimizing the inherent cost of inspection. However, we have already seen in our discussion of specification, that in order to make the greatest number of valid predictions upon the basis of evidence provided by a sample, it is desirable to take into account not only the observed measurements of quality but also the order of these as determined by the sequence in which the pieces of product were turned out. In fact, as has been pointed out elsewhere, if it can be shown that when the production process is in a state of statistical control, the necessary cost of inspection is a minimum.

Four Significant Ways in Which Quality May Differ from Standard

If the consumer is to deal understandingly with the problem of specifying the desired degree of control of variability in quality even under the simplest conditions, it is necessary to take into account at least four of the senses in which observed variability may differ significantly from standard.

I shall assume that the consumer is interested in specifying the quality that the goods are supposed to have after he accepts them. We have seen that it is not possible to write a specification of this kind that can be met in practice until one knows enough about the manufacturing process to justify him in the belief that it is possible to attain the degree of control specified. Let us for our present purpose assume that this knowledge is available and that we are to consider here simply the problem of specifying quality in terms of a set of operations of measurement such as indicated in (2).

We must keep in mind that the consumer must make his choice of goods upon the basis of available evidence that they will prove to be of the quality wanted. It is evident that such a choice must differentiate between the kind of variability specified as standard and that which differs "significantly" from this standard.

The consumer may look for variations in quality beyond a range that would be likely to occur if the product was controlled in the manner called for by the specification.

Such variations may be termed statistically significant. In this way, statistical theory provides us with a means of testing the hypothesis that the quality of a given product is standard.

It is obvious, however, that the practical man is not necessarily interested in the fact that there is an indication that the quality of a given product differs significantly in the statistical sense from standard. That difference, for example, may be so very small that no one could reasonably hope to find and remove the cause of the difference. It is therefore necessary to have some kind of test for causes of variability that we may reasonably hope to find and remove. One such test is provided by the operational techniques of statistical control, in particular the quality control chart. It is of interest to note that the object of detecting such a difference is twofold. First, it indicates a source of lack of homogeneity in the quality of material already produced and, second, it indicates the presence of a cause of variability within the production process that may produce undesirable variations in the quality of future product. In this second sense, the difference is one that points to the desirability of doing something to the production process, provided it is to be used to produce future product of the kind specified.

It is conceivable, however, that the quality of product may differ from standard enough to be detected by either or both a test for statistical significance and a test for lack of control and yet not be large enough to be sensed by the consumer. For example, such differences are usually expressed in terms of physical and chemical measurements and not directly in terms of sensory units. Hence, if the differences are not large enough to be sensed either directly or indirectly by the consumer, they will not be of much interest to him and will not change his valuation of the product from what it would be if it were exactly standard.

However, a difference may be large enough to be sensed and yet not large enough to give the product a wantability which is distinctly less than the wantability of that specified as standard. The fourth kind of significant difference is that which is just large enough to produce a noticeable decrease in the wantability of product. The layman might, of course, jump to the conclusion that this is the only difference of practical significance. To do so would be reasonable only if we had a method of determining with certainty when such differences exist. However, we can never be sure of the existence of such differences. Instead, the existence of such differences can only be inferred in the probability sense upon the basis of available evidence. But if such evidence is to be interpreted with the maximum degree of assurance, it is necessary for us to be pretty sure that assignable causes of variability within the data are absent. Hence in order to determine at reasonable cost whether or not observed differences are significant from the viewpoint of indicating differences large enough to modify wantability, it is necessary first of all to look at the data, particularly from the viewpoint of finding differences that indicate the presence of assignable causes of variability.

CONCLUSION

We are in a position to view in a somewhat more general way the fundamental nature of a standard of quality as an instrument to be used by producers and consumers as a basis for exchange of goods. Since it would be rather useless to specify quality that is not attainable under practical conditions of production, it is not feasible to write a specification in what might be considered permanent form until we have attained knowledge of a high degree of assurance about what is attainable under commercial conditions. However, one cannot wait indefinitely for science and engineering to advance to a state where we have that knowledge. In fact, it is only by setting up a tentative standard, trying it out as it were, modifying it in the light of results, and closely coordinating this modification of the specification with the results of production and inspection that we can hope in the end to approach that knowledge which is necessary before one can set up once and for all the ideal requirements in terms of the maximum degree of control of the variability of quality of manufactured product. In this sense, a tentative specification serves a very useful purpose in the very process of acquiring the knowledge necessary in order to make possible an improved specification. From this viewpoint, a specification serves very much the purpose served by a hypothesis in scientific method.

In conclusion, we may say that a standard of quality is not a written finality but a dynamic process. It is not merely the imprisonment of the past in the form of a specification but also the unfolding of the future as revealed in the steps of production and inspection.