Realistic Evaluation of the Precision and Accuracy of Instrument Calibration Systems *

Churchill Eisenhart

(November 28, 1962)

Calibration of instruments and standards is a refined form of measurement. Measurement of some property of a thing is an operation that yields as an end result a number that indicates how much of the property the thing has. Measurement is ordinarily a repeatable operation, so that it is appropriate to regard measurement as a production process, the "product" being the numbers, i.e., the measurements, that it yields; and to apply to measurement processes in the laboratory the concepts and techniques of statistical process control that have proved so useful in the quality control of industrial production.

Viewed thus it becomes evident that a particular measurement operation cannot be regarded as constituting a measurement process unless statistical stability of the type known as a state of statistical control has been attained. In order to determine whether a particular measurement operation is, or is not, in a state of statistical control it is necessary to be definite on what variations of procedure, apparatus, environmental conditions, observers, operators, etc., are allowable in "repeated applications" of what will be considered to be the same measurement process applied to the measurement of the same quantity under the same conditions. To be realistic, the "allowable variations" must be of sufficient scope to bracket the circumstances likely to be met in practice. Furthermore, any experimental program that aims to determine the standard deviation of a measurement process as an indication of its precision, must be based on appropriate random sampling of this likely range of circumstances.

Ordinarily the accuracy of a measurement process may be characterized by giving (a) the standard deviation of the process and (b) credible bounds to its likely overall systematic error. Determination of credible bounds to the combined effect of recognized potential sources of systematic error always involves some arbitrariness, not only in the placing of reasonable bounds on the systematic error likely to be contributed by each particular assignable cause, but also in the manner in which these individual contributions are combined. Consequently, the "inaccuracy" of end results of measurement cannot be expressed by "confidence limits" corresponding to a definite numerical "confidence level," except in those rare instances in which the possible overall systematic error of a final result is negligible in comparison with its imprecision.

1. Introduction

Calibration of instruments and standards is basically a refined form of measurement. Measurement is the assignment of numbers to material things to represent the relations existing among them with respect to particular properties. One always measures properties of things, not the things themselves. In practice, measurement of some property of a thing ordinarily takes the form of a sequence of steps or operations that yields as an end result a number that indicates how much of this property the thing has, for someone to use for a specific purpose. The end result may be the outcome of a single reading of an instrument. More often it is some kind of average, e.g., the arithmetic mean of a number of independent determinations of the same magnitude, or the final result of a least squares "reduction" of measurements of a number of different quantities that bear known relations to each other in accordance with a definite experimental plan. In general, the purpose for which the answer is needed determines the accuracy required and ordinarily also the method of measurement employed.

Specification of the apparatus and auxiliary equipment to be used, the operations to be performed, the sequence in which they are to be executed, and the conditions under which they are respectively to be carried out—these instructions collectively serve to define a method of measurement. A measurement process is the realization of a method of measurement in terms of particular apparatus and equipment of the prescribed kinds, particular conditions that at best only approximate the conditions prescribed, and particular persons as operators and observers.

It has long been recognized that, in undertaking to apply a particular method of measurement, a degree of consistency among repeated measurements of a single quantity needs to be attained before the method of measurement concerned can be regarded as meaningfully realized, i.e., before a measurement process can be said to have been established that is

a realization of the method of measurement concerned. Indeed, consistency or statistical stability of a very special kind is required: to qualify as a measurement process a measurement operation must have attained what is known in industrial quality control language as a state of statistical control. Until a measurement operation has been "debugged" to the extent that it has attained a state of statistical control it cannot be regarded in any logical sense as measuring anything at all. And when it has attained a state of statistical control there may still remain the question of whether it is faithful to the method of measurement of which it is intended to be a realization.

The systematic error, or bias, of a measurement process refers to its tendency to measure something other than what was intended; and is determined by the magnitude of the difference $\mu - \tau$ between the process average or limiting mean μ associated with measurement of a particular quantity by the measurement process concerned and the true value τ of the magnitude of this quantity. On first thought, the "true value" of the magnitude of a particular quantity appears to be a simple straightforward concept. On careful analysis, however, it becomes evident that the "true value" of the magnitude of a quantity is intimately linked to the purposes for which knowledge of the magnitude of this quantity is needed, and cannot, in the final analysis, be meaningfully and usefully defined in isolation from these needs.

The precision of a measurement process refers to, and is determined by the degree of mutual agreement characteristic of independent measurements of a single quantity yielded by repeated applications of the process under specified conditions; and its accuracy refers to, and is determined by, the degree of agreement of such measurements with the true value of the magnitude of the quantity concerned. In brief "accuracy" has to do with closeness to the truth; "precision," only with closeness together.

Systematic error, precision, and accuracy are inherent characteristics of a measurement process and not of a particular measurement yielded by the process. We may also speak of the systematic error, precision, and accuracy of a particular method of measurement that has the capability of statistical control. But these terms are not defined for a measurement operation that is not in a state of statistical control.

The precision, or more correctly, the *imprecision* of a measurement process is ordinarily summarized by the standard deviation of the process, which expresses the characteristic disagreement of repeated measurements of a single quantity by the process concerned, and thus serves to indicate by how much a particular measurement is likely to differ from other values that the same measurement process might have provided in this instance, or might yield on remeasurement of the same quantity on another occasion. Unfortunately, there does not exist any single comprehensive measure of the accuracy (or *inaccuracy*) of a measurement process analogous to the standard deviation as a measure of its imprecision.

To characterize the accuracy of a measurement process it is necessary, therefore, to indicate (a) its systematic error or bias, (b) its precision (or *imprecision*)—and, strictly speaking, also, (c) the form of the distribution of the individual measurements about the process average. Such is the unavoidable situation if one is to concern one's self with individual measurements yielded by any particular measurement process. Fortunately, however, "final results" are ordinarily some kind of average or adjusted value derived from a set of independent measurements, and when four or more independent measurements are involved, such adjusted values tend to be normally distributed to a very good approximation, so that the accuracy of such final results can ordinarily be characterized satisfactorily by indicating (a) their imprecision as expressed by their standard error, and (b) the systematic error of the process by which they were obtained.

The error of any single measurement or adjusted value of a particular quantity is, by definition, the difference between the measurement or adjusted value concerned and the true value of the magnitude of this quantity. The error of any particular measurement or adjusted value is, therefore, a fixed number; and this number will ordinarily be unknown and unknowable, because the true value of the magnitude of the quantity concerned is ordinarily unknown and unknowable. Limits to the error of a single measurement or adjusted value may, however, be inferred from (a) the precision, and (b) bounds on the systematic error of the measurement process by which it was produced—but not without risk of being incorrect, because, quite apart from the inexactness with which bounds are commonly placed on a systematic error of a measurement process, such limits are applicable to the error of the single measurement or adjusted value, not as a unique individual outcome, but only as a typical case of the errors characteristic of such measurements of the same quantity that might have been, or might be, yielded by the same measurement process under the same conditions.

Since the precision of a measurement process is determined by the characteristic "closeness together" of successive independent measurements of a single magnitude generated by repeated application of the process under specified conditions, and its bias or systematic error is determined by the direction and amount by which such measurements tend to differ from the true value of the magnitude of the quantity concerned, it is necessary to be clear on what variations of procedure, apparatus, environmental conditions, observers, etc., are allowable in "repeated applications" or what will be considered to be the same measurement process applied to the measurement of the same quantity under the same conditions. If whatever measures of the precision and bias of a measurement process we may adopt are to provide a realistic indication of the accuracy of this process in practice, then the "allowable variations" must be of sufficient scope to bracket the range of circumstances commonly met in practice. Furthermore, any experimental program that aims to determine the pre-

162

cision, and thence the accuracy of a measurement process, must be based on an appropriate random sampling of this "range of circumstances," if the usual tools of statistical analysis are to be strictly applicable.

When adequate random sampling of the appropriate "range of circumstances" is not feasible, or even possible, then it is necessary (a) to compute, by extrapolation from available data, a more or less subjective estimate of the precision of the measurement process concerned, to serve as a substitute for a direct experimental measure of this characteristic, and (b) to assign more or less subjective bounds to the systematic error of the measurement process. To the extent that such at least partially subjective computations are involved, the resulting evaluation of the overall accuracy of a measurement process "is based on subject-matter knowledge and skill, general information, and intuition—but not on statistical methodology" [Cochran et al. 1953, p. 693]. Consequently, in such cases the statistically precise concept of a family of "confidence intervals" associated with a definite "confidence level" or "confidence coefficient" is not applicable.

The foregoing points and certain other related matters are discussed in greater detail in the succeeding sections, together with an indication of procedures for the realistic evaluation of precision and accuracy of established procedures for the calibration of instruments and standards that minimize as much as possible the subjective elements of such an evaluation. To the extent that complete elimination of the subjective element is not always possible, the responsibility for an important and sometimes the most difficult part of the evaluation is shifted from the shoulders of the statistician to the shoulders of the subject matter "expert."

2. Measurement

2.1. Nature and Object

Measurement is the assignment of numbers to material things to represent the relations existing among them with respect to particular properties. The number assigned to some particular property serves to represent the relative amount of this property associated with the object concerned.

Measurement always pertains to properties of things, not to the things themselves. Thus we cannot measure a meter bar, but can and usually do, measure its length; and we could also measure its mass, its density, and perhaps, also its hardness.

The object of measurement is twofold: first, symbolic representation of properties of things as a basis for conceptual analysis; and second, to effect the representation in a form amenable to the powerful tools of mathematical analysis. The decisive feature is symbolic representation of properties, for which end numerals are not the only usable symbols.

In practice the assignment of a numerical magnitude to a particular property of a thing is ordinarily accomplished by comparison with a set of standards, or by comparison either of the quantity itself, or of some transform of it, with a previously calibrated scale. Thus, length measurements are usually made by directly comparing the length concerned with a calibrated bar or tape; and mass measurements, by directly comparing the weight of a given mass with the weight of a set of standard masses, by means of a balance; but force measurements are usually carried out in terms of some transform, such as by reading on a calibrated scale the extension that the force produces in a spring, or the deflection that it produces in a proving ring; and temperature measurements are usually performed in terms of some transform, such as by reading on a calibrated scale the expansion of a column of mercury, or the electrical resistance of a platinum wire.

2.2. Qualitative and Quantitative Aspects

As Walter A. Shewhart, father of statistical control charts, has remarked:

"It is important to realize . . . that there are two aspects of an operation of measurement; one is quantitative and the other qualitative. One consists of *numbers* or pointer readings such as the observed lengths in n measurements of the length of a line, and the other consists of the *physical manipulations* of physical things by *someone* in accord with instructions that we shall assume to be describable in words constituting a text." [Shewhart 1939, p. 130.]

More specifically, the qualitative factors involved in the measurement of a quantity are: the *apparatus* and *auxiliary equipment* (e.g., reagents, batteries or other source of electrical energy, etc.) employed; the *operators* and *observers*, if any, involved; the *operations* performed, together with the *sequence* in which, and the *conditions* under which, they are respectively carried out.

2.3. Correction and Adjustment of Observations

The numbers obtained as "readings" on a calibrated scale are ordinarily the end product of everyday measurement in the trades and in the home. In scientific work there are usually two important additional quantitative aspects of measurement: (1) *correction* of the readings, or their transforms, to compensate for known deviations from ideal execution of the prescribed operations, and for nonnegligible effects of variations in uncontrolled variables; and (2) *adjustment* of "raw" or corrected measurements of particular quantities to obtain values of these quantities that conform to restrictions upon, or interrelations among, the magnitudes of these quantities imposed by the nature of the problem.

Thus, it may not be practicable or economically feasible to take readings at exactly the prescribed temperatures; but quite practicable and feasible to bring and hold the temperature within narrow neighborhoods of the prescribed values and to record the actual temperatures to which the respective readings correspond. In such cases, if the deviations from the prescribed temperatures are not negligible, "temperature *corrections*" based on appropriate theory are usually applied to the respective readings to bring

them to the values that presumably would have been observed if the temperature in each instance had been exactly as prescribed.

In practice, however, the objective just stated is rarely, if ever, actually achieved. Any "temperature corrections" applied could be expected to bring the respective readings "to the values that presumably would have been observed if the temperature in each instance had been exactly as prescribed" if and only if these "temperature corrections" made appropriate allowances for *all* of the effects of the deviations of the actual temperatures from those prescribed. "Temperature corrections" ordinarily correct *only* for particular effects of the deviations of the actual temperatures from their prescribed values; *not* for all of the effects on the readings traceable to deviations of the actual temperatures from those prescribed. Thus Michelson utilized *"temperature corrections"* in his 1879 investigation of the speed of light; but his results exhibit a dependence on temperature after "temperature correction." The "temperature corrections" applied corrected only for the effects of thermal expansion due to variations in temperature and not also for changes in the index of refraction of the air due to changes in the humidity of the air, which in June and July at Annapolis is highly correlated with temperature. *Corrections applied in practice are usually of more limited scope than the names that they are given appear to indicate.*

Adjustment of observations is fundamentally different from their "correction." When two or more related quantities are measured individually, the resulting measured values usually fail to satisfy the constraints on their magnitudes implied by the given interrelations among the quantities concerned. In such cases these "raw" measured values are mutually contradictory, and require *adjustment* in order to be usable for the purpose intended. Thus, measured values of the three cyclic differences $(A-B)$, $(B-C)$, and $(C-A)$ between the lengths of three nominally equivalent gage blocks are mutually contradictory, and strictly speaking are not usable as values of these differences, unless they sum to zero.

The primary goal of *adjustment* is to derive from such inconsistent measurements, if possible, *adjusted values* for the quantities concerned that do satisfy the constraints on their magnitudes imposed by the nature of the quantities themselves and by the existing interrelations among them. A second objective is to select from all possible sets of adjusted values the set that is the "best"—or, at least, a set that is "good enough" for the intended purpose—in some well-defined sense. Thus, in the above case of the measured differences between the lengths of three gage blocks, an adjustment could be effected by ignoring the measured value of one of the differences entirely, say, the difference $(C-A)$, and taking the negative of the sum of the other two as its adjusted value,

$$Adj(C-A) = -[(A-B)+(B-C)].$$

This will certainly assure that the sum of all three values, $(A-B)+(B-C)+Adj(C-A)$, is zero, as required, and is clearly equivalent to ascribing all of the excess or deficit to the replaced measurement, $(C-A)$. Alternatively, one might prefer to distribute the necessary total adjustment $-[(A-B)+(B-C)+(C-A)]$ equally over the individual measured differences, to obtain the following set of adjusted values:

$$Adj(A-B) = (A-B) - \frac{1}{3}[(A-B)+(B-C)+(C-A)]$$

$$= \frac{1}{3}[2(A-B)-(B-C)-(C-A)]$$

$$Adj(B-C) = \frac{1}{3}[2(B-C)-(A-B)-(C-A)]$$

$$Adj(C-A) = \frac{1}{3}[2(C-A)-(A-B)-(B-C)].$$

Clearly, the sum of these three adjusted values must always be zero, as required, regardless of the values of the original individual measured differences. Furthermore, most persons, I believe, would consider this latter adjustment the better; and under certain conditions with respect to the "law of error" governing the original measured differences, it is indeed the "best."

Note that no adjustment problem existed at the stage when only two of these differences had been measured whichever they were, for then the third could be obtained by subtraction. As a general principle, when no more observations are taken than are sufficient to provide one value of each of the unknown quantities involved, then the results so obtained are usable at least—they may not be "best." On the other hand, when additional observations are taken, leading to "over determination" and consequent contradiction of the fundamental properties of, or the basic relationships among the quantities concerned, then the respective observations must be regarded as contradicting one another. When this happens the observations themselves, or values derived from them, must be replaced by adjusted values such that all contradiction is removed. "This is a logical necessity, since we cannot accept for truth that which is contradictory or leads to contradictory results." [Chauvenet 1868, p. 472.]

2.4. Scheduling the Taking of Measurements

Having done what one can to remove extraneous sources of error, and to make the basic measurements as precise and as free from systematic error as possible, it is frequently possible not only to increase the precision of the end results of major interest but also to simultaneously decrease their sensitivity to sources of possible systematic error, by careful scheduling of the measurements required. An instance is provided by the traditional procedure for calibrating liquid-in-glass thermometers [Waidner and Dickinson 1907, p. 702; NPL 1957, pp. 29–30; Swindells 1959, pp. 11–12]: Instead of attempting to hold the temperature of the comparison bath constant, a very difficult objective to achieve, the heat

input to the bath is so adjusted that its temperature is slowly increasing at a steady rate, and then readings of, say, four test thermometers and two standards are taken in accordance with the schedule

$$S_1 T_1 T_2 T_3 T_4 S_2 S_2 T_4 T_3 T_2 T_1 S_1$$

the readings being spaced uniformly in time so that the arithmetic mean of the two readings of any one thermometer will correspond to the temperature of the comparison bath at the midpoint of the period. Such scheduling of measurement taking operations so that the effects of the specific types of departures from perfect control of conditions and procedure will have an opportunity to balance out is one of the principal aims of the art and science of *statistical design of experiments*. For additional physical science examples, see, for instance, Youden [1951a; and 1954–1959].

2.5. Measurement as a Production Process

We may summarize our discussion of measurement up to this point, as follows: Measurement of some property of a thing in practice always takes the form of a sequence of steps or operations that yield as an end result a number that serves to represent the amount or quantity of some particular property of a thing—a number that indicates how much of this property the thing has, for someone to use for a specific purpose. The end result may be the outcome of a single reading of an instrument, with or without corrections for departures from prescribed conditions. More often it is some kind of average or adjusted value, e.g., the arithmetic mean of a number of independent determinations of the same magnitude, or the final result of, say, a least squares "reduction" of measurements of a number of different quantities that have known relations to the quantity of interest.

Measurement of some property of a thing is ordinarily a repeatable operation. This is certainly the case for the types of measurement ordinarily met in the calibration of standards and instruments. It is instructive, therefore, to regard measurement as a *production process*, the "product" being the numbers, that is, the measurements that it yields; and to compare and contrast measurement processes in the laboratory with mass production processes in industry. For the moment it will suffice to note (a) that when successive amounts of units of "raw material" are processed by a particular mass production process, the output is a series of nominally identical items of product—of the particular type produced by the mass production operation, i.e., by the *method of production* concerned; and (b) that when successive objects are measured by a particular measurement process, the individual items of "product" produced consist of the numbers assigned to the respective objects to represent the relative amounts that they possess of the property determined by the *method of measurement* involved.

2.6. Methods of Measurement and Measurement Processes

Specification of the apparatus and auxiliary equipment to be used, the operations to be performed, the sequence in which they are to be carried out, and the conditions under which they are respectively to be carried out—these *instructions* collectively serve to define a *method of measurement*. To the extent that corrections may be required they are an integral part of measurement. The types of corrections that will ordinarily need to be made, and specific procedures for making them, should be included among "the operations to be performed." Likewise, the essential adjustments required should be noted, and specific procedures for making them incorporated in the specification of a method of measurement.

A *measurement process* is the realization of a method of measurement in terms of particular apparatus and equipment of the prescribed kinds, particular conditions that at best only approximate the conditions prescribed, and particular persons as operators and observers [ASTM 1961, p. 1758; Murphy 1961, p. 264]. Of course, there will often be a question whether a particular measurement process is loyal to the method of measurement of which it is intended to be a realization; or whether two different measurement processes can be considered to be realizations of the same method of measurement.

To begin with, written specifications of methods of measurement often contain absolutely precise instructions which, however, cannot be carried out (repeatedly) with complete exactitude in practice; for example, "move the two parallel cross hairs of the micrometer of the microscope until the graduation line of the standard is centered between them." The accuracy with which such instructions can be carried out in practice will always depend upon "the circumstances"; in the case cited, on the skill of the operator, the quality of the graduation line of the standard, the quality of the screw of the micrometer, the parallelism of the cross hairs, etc. To the extent that the written specification of a method of measurement involves absolutely precise instructions that cannot be carried out with complete exactitude in practice there are certain to be discrepancies between a method of measurement and its realization by a particular measurement process.

In addition, the specification of a method of measurement often includes a number of imprecise instructions, such as "raise the temperature slowly," "stir well before taking a reading," "make sure that the tubing is clean," etc. Not only are such instructions inherently vague, but also in any given instance they must be understood in terms of the general level of refinement characteristic of the context in which they occur. Thus, "make sure that the tubing is clean" is not an absolutely definite instruction; to some people this would mean simply that the tubing should be clean enough to drink liquids through; in some laboratory work it might be interpreted to mean mechanically washed and scoured so as to be free from dirt and other ordinary

solid matter (but not cleansed also with chemical solvents to remove more stubborn contaminants); to an advanced experimental physicist it may mean not merely mechanically washed and chemically cleansed, but also "out gassed" by being heated to and held at a high temperature, near the softening point, for an hour or so. All will agree, I believe, that it would be exceedingly difficult to make such instructions absolutely definite with a convenient number of words. To the extent that the specification of a method of measurement includes instructions that are not absolutely definite, there will be room for differences between measurement processes that are intended to be realization of the very same method of measurement.

Recognition of the difficulty of achieving absolute definiteness in the specification of a method of measurement does not imply that "any old set" of instructions will serve to define a method of measurement. Quite the contrary. To qualify as a specification of a method of measurement, a set of instructions must be sufficiently definite to insure statistical stability of repeated measurements of a single quantity, that is, derived measurement processes must be capable of meeting the criteria of *statistical control* [Shewhart 1939, p. 131; Murphy 1961, p. 265; ASTM 1961, p. 1758]. To elucidation of the meaning of, and need for this requirement we now turn.

3. Properties of Measurement Processes

3.1. Requirement of Statistical Control

The need for attaining a degree of consistency among repeated measurements of a single quantity before the method of measurement concerned can be regarded as meaningful has certainly been recognized for a long, long time. Thus Galileo, describing his famous experiment on the acceleration of gravity in which he allowed a ball to roll different distances down an inclined plane wrote:

". . . si lasciava (como dico) scendere per il detto canale la palla, notando, nel modo che appresso dirò, il temp che consumava nello scorrerlo tutto, replicando il medesimo atto molte volte per assicurarsi bene della quantità del temp, nel quale non si trovava mai differenza nè anco della decima parte d'una battuta di polso. Fatta e stabilita precisamente tale operazione, facemmo scender la medisima palla solamente per la quarta parte della lunghezza di esso canale . . ." [1] [Galileo 1638, Third Day; Nat'l. ed., p. 213.]

Something more than mere "consistency" is required, however, as Shewhart points out eloquently in his very important chapter on "The Specification of Accuracy and Precision" [Shewhart 1939, ch. IV]. He begins by noting that the description given by R. A. Millikan [1903, pp. 195–196] of a method for determining the surface tension T of a liquid from measurements of the force of tension F of a film of the liquid contains the following instruction with regard to the basic readings from which measurements of F are derived: "Continue this operation until a number of consistent readings can be obtained." Shewhart then comments on this as follows:

". . . the text describing the operation does not say to carry out such and such physical operations and call the result a measurement of T. Instead, it says in effect not to call the result a measurement of T until one has attained a certain degree of *consistency* among the observed values of F and hence among those of T. Although this requirement is not always explicitly stated in specifications of the operation of measurements as it was here, I think it is always implied. Likewise, I think it is always assumed that there can be too much consistency or uniformity among the observed values as, for example, if a large number of measurements of the surface tension of a liquid were found to be identical. What is wanted but not explicitly described is a specific kind and degree of consistency.

". . . it should be noted that the advice to repeat the operation of measuring surface tension until a number of consistent readings have been obtained is indefinite in that it does not indicate how many readings shall be taken before applying a test for consistency, nor what kind of test of consistency is to be applied to the numbers or pointer readings . . . . One of the objects of this chapter is to see how far one can go toward improving this situation by providing an operationally definite criterion that preliminary observations must meet before they are to be considered consistent in the sense implied in the instruction cited above.

"Before doing this, however, we must give attention not so much to the consistency of the n observed values already obtained by n repetitions of the operation of measurement as we do to the *reproducibility of the operation* as determined by the numbers in the potentially infinite sequence corresponding to an infinite number of repetitions of this operation. No one would care very much how consistent the first n preliminary observations were if nothing could be validly inferred from this as to what future observations would show. Hence, it seems to me that the characteristics of the numerical aspects of an operation that is of greatest practical interest is its *reproducibility within tolerance limits throughout the infinite sequence*. The limit to which we may go in this direction is to attain a state of statistical control. The attempt to attain a certain kind of consistency within the first n observed values is merely a means of attaining reproducibility within limits throughout the whole of the sequence." [Shewhart 1939, pp. 131–132.]

The point that Shewhart makes forcefully, and stresses repeatedly later in the same chapter, is that the first n measurements of a given quantity generated by a particular measurement process provide a logical basis for predicting the behavior of further measurements of the same quantity by the same measurement process if and only if these n measurements may be regarded as a *random sample* from a "population" or "universe" of all conceivable measurements of the given quantity by the measurement process concerned; that is, in the language of mathematical statistics, if an only if the n measurements in hand may be regarded as "observed values" of a sequence of random variables characterized by a probability distribution identified with the measurement process concerned, and related through the values of one or more of its parameters to the magnitude of the quantity measured.

It should be noted especially that nothing is said about the mathematical form of the *probability distribution* of these random variables. The important thing is that there be one. W. Edwards

[1] I am grateful to my colleague Ugo Fano for the following literal translation: ". . . we let, as I was saying, the ball descend through said channel, recording, in a manner presently to be described, the time it took in traversing it all, repeating the same action many times to make really sure of the magnitude of time, in which one never found a difference of even a tenth of a pulsebeat. Having done and established precisely such operation, we let the same ball descend only for the fourth part of the length of the same channel; . . ."

Deming has put this clearly and forcefully in these words:

"In applying statistical theory, the main consideration is not what the shape of the universe is, but whether there is any universe at all. No universe can be assumed, nor . . . statistical theory . . . applied unless the observations show statistical control. In this state the samples when cumulated over a suitable interval of time give a distribution of a particular shape, and this shape is reproduced hour after hour, day after day, so long as the process remains in statistical control—i.e., exhibits the properties of randomness. In a state of control, n observations may be regarded as a sample from the universe of whatever shape it is. A big enough sample, or enough small samples, enables the statistician to make meaningful and useful predictions about future samples. This is as much as statistical theory can do.

". . . Very often the experimenter, instead of rushing in to apply [statistical methods] should be more concerned about attaining statistical control and asking himself whether any predictions at all (the only purpose of his experiment), by statistical theory or otherwise, can be made." [Deming 1950, pp. 502–503.]

Shewhart was well aware of the fact that from a set of n measurements in hand it is not possible to decide with absolute certainty whether they do or do not constitute a *random sample* from some definite statistical "population" characterized by a probability distribution. He, therefore, proposed [Shewhart 1939, pp. 146–147] that in any particular instance one should "decide to act for the present as if"[2] the measurements in hand (and their immediate successors) were a simple random sample from a definite statistical population—i.e., in the language of mathematical statistics, were "observed values" of *independent identically distributed random variables*—only if the measurements in hand met the requirements of the small-samples version of Criterion I of his previous book [Shewhart 1931, pp. 309–318] and of certain additional tests of randomness that he described explicitly for the first time in his contribution to the University of Pennsylvania Bicentennial Conference in September 1940 [Shewhart, 1941]. In other words, Shewhart proposed that one should consider a measurement process to be—i.e., should "decide to act for the present as if" the process were—in a *state of (simple) statistical control*, only if the measurements in hand show no evidence of lack of statistical control when analyzed for randomness *in the order in which they were taken* by the control chart techniques for averages and standard deviations that he had found so valuable in industrial process control and by certain additional tests for randomness based on "runs above and below average" and "runs up and down."[3]

Simpson[3] did not prove that taking of the Arithmetic Mean was the best thing to do but merely that it is good. However, in accomplishing this goal he did something much more important: he took the bold step of regarding errors of measurement, not as unique unrelated magnitudes unamenable to mathematical analysis, but as distributed in accordance with a probability distribution that was an intrinsic property of the measurement process itself. He thus opened the way to a mathematical theory of measurement based on the mathematical theory of probability; and, in particular, to the formulation and development of the Method of Least Squares in essentially its present day form by Gauss (1809, 1821) and Laplace (1812).

"Student" (William Sealy Gosset, 1876–1937), pioneer statistical consultant and "father" of the "theory of small samples," was certainly among the first to stress the importance of randomness in measurement and experimentation. Thus, he began his revolutionary 1908 paper on "The probable error of a mean" with these remarks:

"Any experiment may be regarded as forming an individual of a 'population' of experiments which might be performed under the same conditions. A series of experiments is a sample drawn from this population.

"Now any series of experiments is only of value in so far as it enables us to form a judgment as to the statistical constants of the population to which the experiments belong." [Student 1908, p. 1.]

None of these writers, nor any of their contemporaries, however, provided "an operationally definite criterion that preliminary observations must meet" before we take it upon ourselves "to act for the present as if" they and their immediate successors were random samples from a "population" or "universe" of all conceivable measurements of the given quantity by the measurement process concerned. Provision of such a criterion is Shewhart's major contribution.

Experience shows that in the case of measurement processes the ideal of strict statistical control that Shewhart prescribes is usually very difficult to attain, just as in the case of industrial production processes. Indeed, many measurement processes simply do not and, it would seem, cannot be made to conform to this ideal of producing successive measurements of a single quantity that can be considered to be "observed values" of independent identically distributed random variables.[4] The nature of the "trouble" was stated succinctly by Student in 1917 when, speaking of physical and chemical determinations, he wrote:

"After considerable experience I have not encountered any determination which is not influenced by the date on which it is made; from this it follows that a number of determinations of the same thing made on the same day are likely

---

[2] This very explicit phraseology is due to John W. Tukey [1960, p. 424].

[3] Thomas Simpson, in his now famous letter [Simpson 1755] to the President of the Royal Society of London "on the Advantage of taking the Mean of a Number of Observations, in practical Astronomy," was the first to consider repeated measurements of a single quantity by a given measurement process as observed values of independent random variables having the same probability distribution. His conclusion is of interest in itself:

"Upon the whole of which it appears, that the taking of the Mean of a number of observations, greatly diminishes the chances for all the smaller errors, and cuts off almost all possibility of any great ones: which last consideration, alone, seem sufficient to recommend the use of the method, not only to astronomers, but to all others concerned in making of experiments of any kind (to which the above reasoning is equally applicable). And the more observations or experiments there are made, the less will the conclusion be liable to err, provided they admit of being repeated under the same circumstances."

---

[4] Looking at the matter from a fundamental viewpoint, perhaps we should say, not that Shewhart's ideal of strict statistical control is unattainable in the case of such measurement processes, but rather that the degree of approximation to this ideal can be made as close as one chooses, if one is willing to pay the price. In other words, how close one chooses to bring a measurement process to the ideal of strict statistical control is, in any given instance, basically an economic matter, taking into account, of course, not only the immediate purpose(s) for which the measurements are intended but also the other uses to which they may be put. (Compare Simon [1946, p. 566] and Eisenhart [1952, p. 554]).

to lie more closely together than if the repetitions had been made on different days." [Student 1917, p. 415.]

In other words, production of measurements seems to be like the production of paint; and just as in the case of paint, if one must cover a large surface all of which is visible simultaneously, one will do well to use paint all from the same batch, so in the case of measurements, if a scientist or metrologist "wishes to impress his clients" he will "arrange to do repetition analyses as nearly as possible at the same time." [Student 1927, p. 155.]

Fortunately, just as one may blend paint from several batches to obtain a more uniform color, and one which is, presumably, closer to the "process average," so also may a scientist or metrologist "if he wishes to diminish his real error, . . . separate [his measurements] by as wide an interval of time as possible" [Student, loc. cit.] and then take an appropriate average of them as his determination. Consequently, if we are to permit such averaging as an allowable step in a fully specified measurement process (see sec. 2.6 above), then we are obliged to recognize both within-day and between-day components of variation, and accept such a complex measurement process as being in a state of statistical control overall, or as we shall say, in a *state of COMPLEX statistical control*, when the components of within-day and between-day variation are both in a state of statistical control in Shewhart's strict sense, which we shall term *SIMPLE statistical control*. In more complex situations, one may be obliged to recognize more than two "layers" of variation, and, sometimes, more than a single component of variation within a given "layer."

Adopting this more general concept of statistical control, R. B. Murphy of the Bell Telephone Laboratories in his essay "On the Meaning of Precision and Accuracy" [Murphy 1961], published in advance of the issuance by the American Society for Testing and Materials of its Tentative Recommended Practice with respect to the "Use of the Terms Precision and Accuracy as Applied to Measurement of a Property of a Material" [ASTM 1961], remarks:

"Following through with this line of thought borrowed from quality control, we shall add a requirement that an effort to follow a test method ought not to be known as a measurement process unless it is capable of statistical control. Capability of control means that either the measurements are the product of an identifiable statistical universe or an orderly array of such universes or, if not, the physical causes preventing such identification may themselves be identified and, if desired, isolated and suppressed. Incapability of control implies that the results of measurement are not to be trusted as indications of the physical property at hand—in short, we are not in any verifiable sense measuring anything . . . . Without this limitation on the notion of measurement process, one is unable to go on to give meaning to those statistical measures which are basic to any discussion of precision and accuracy." [Murphy 1961, pp. 264–265.]

3.2. Postulate of Measurement and the Concept of a Limiting Mean

A conspicuous characteristic of measurement is disagreement of repeated measurements of the same quantity. Experience shows that, when high accuracy is sought, repeated measurements of the same quantity by a particular measurement process does not yield uniformly the same number.[5] We explain these discordances by saying that the individual measurements are affected by *errors*, which we interpret to be the manifestations of variations in the execution of the process of measurement resulting from "the imperfections of instruments, and of organs of sense," and from the difficulty of achieving (or even specifying with a convenient number of words) the ideal of perfect control of conditions and procedure.

This "cussedness of measurements" brings us face to face with a fundamental question: In what sense can we say that the measurements yielded by a particular measurement process serve to determine a unique magnitude, when experience shows that repeated measurement of a single quantity by this process yields a sequence of nonidentical numbers. What is the value thus determined?

The answer takes the form of a *postulate* about measurement processes that has been expressed by N. Ernest Dorsey, as follows:

"The mean of a family of measurements—of a number of measurements for a given quantity carried out by the same apparatus, procedure and observer—approaches a definite value as the number of measurements is indefinitely increased. Otherwise, they could not properly be called measurements of a given quantity. In the theory of errors, this limiting mean is frequently called the 'true' value, although it bears no necessary relation to the true quaesitum, to the actual value of the quantity that the observer desires to measure. This has often confused the unwary. Let us call it the limiting mean." [Dorsey 1944, p. 4; Dorsey and Eisenhart 1953, p. 103.]

In my lectures at the National Bureau of Standards, and elsewhere, I have termed this—or rather a slightly rephrased version of it—the *Postulate of Measurement*. A mathematical basis for it is provided by the Strong Law of Large Numbers, a theorem in the mathematical theory of probability discovered during the present century. See, for example, Feller [1957, pp. 243–245, 374], Gnedenko [1962, pp. 241–249], or Parzen [1960, p. 420].

Needless to say, by a "family of measurements" Dorsey means, not a succession of "raw" readings, but rather a succession of adjusted or corrected values which, by virtue of adjustment or correction, can rightfully be considered to be determinations of a single magnitude.

a. Mathematical Formulation

The foregoing can be expressed mathematically as follows: on some particular occasion, say the ith, we may take a number of successive measurements of a single quantity by a given measurement process under certain specified circumstances. Let

$$x_{i1},\ x_{i2},\ \ldots,\ x_{ij},\ \ldots \tag{1}$$

denote the sequence of measurements so generated. Conceptually at least, this sequence could be continued indefinitely. Likewise, on different occasions we might start a new sequence, using the same measurement procedure and applying it to measurement of the same quantity under the same fixed set of circumstances. Each such fresh "start" would correspond to a different value of i. If, for example, the measurement process concerned is statistically stable in the sense of being in a *state of statistical control* as defined by Shewhart [1939], then the Strong Law of Large Numbers will be applicable and we may expect the sequence of cumulative arithmetic means on the ith occasion, namely,

$$\overline{x}_{in} \equiv (x_{i1} + x_{i2} + \ldots + x_{in})/n, \quad (n = 1, 2, \ldots), \quad (2)$$

to converge to μ, a number that constitutes the limiting mean associated with the quantity measured by this measurement process under the circumstances concerned, but independent of the "occasion," that is, independent of the value of "i." The Strong Law of Large Numbers does not guarantee that the sequence (2) for a particular value of "i" will converge to μ as the number of observations n on this occasion tends to infinity, but simply states that among the family of such sequences corresponding to a large number of different starts, ($i = 1, 2, \ldots$), the *instances of nonconvergence to μ* will be *rare exceptions*. In other words, if the measurement process with which one is concerned satisfies the conditions for validity of the Strong Law of Large Numbers, then in practice one is almost certain to be working with a "good" sequence—one for which (2) would converge to μ if the number of observations were continued indefinitely—but "bad" occasions can occur, though rarely. Thus, the Postulate of Measurement expresses something better than an "on-the-average" property—it expresses an "in-almost-all-cases" property. Furthermore, this limiting mean μ, the value of which each individual measurement x is trying to express, can be regarded not only as the *mean* or "center of gravity" of the infinite conceptual population of all measurements x that might conceivably be generated by the measurement process concerned under the specified circumstances, but also as *the* value of the quantity concerned as determined by *this* measurement process.

b. Aim of the Postulate

The sole aim of the Postulate of Measurement is axiomatic acceptance of the existence of a limit approached by the arithmetic mean of a finite number n of measurements generated by any measurement process as $n \to \infty$. It says nothing about how the "best" estimate of this limiting mean is to be obtained from a finite number of such observations. The Postulate is an answer to the need of the practical man for a justification of his desire to consider the sequence of nonidentical numbers that he obtains when he attempts to measure a quantity "by the same method under like circumstances" as pertaining to a single magnitude, in spite of the evident discordance of its elements. The Postulate aims to satisfy this need by telling him that if he were to continue taking more and still more measurements on this quantity "by the same method under like circumstances" ad infinitum, and were to calculate their cumulative arithmetic means at successive stages of this undertaking, then he would find that the successive terms of this sequence of cumulative arithmetic means would settle down to a narrower and ever narrower neighborhood of some definite number which he could then accept as *the* value of the magnitude that his first few measurements were striving to express.

c. Importance of Limiting Mean

The concept of a *limiting mean* associated with the measurement of a given quantity by a particular measurement process that is in a *state of statistical control* is important because by means of statistical methods based on the mathematical theory of probability we can make quantitative inferential statements, with known chances of error, about the magnitude of this limiting mean from a set of measurements of the given quantity by the measurement process concerned. The magnitude of the limiting mean associated with the measurement of a given quantity by a particular measurement process must be carefully distinguished from the *true magnitude* of the quantity measured, about which we may be tempted to make similar inferential statements. Insofar as we make statistical inferences from a set of measurements, we make them with respect to a property of the measurement process involved under the circumstances concerned. The step from quantitative inferential statements about the limiting mean associated with the measurement of a given quantity by a particular measurement process, to quantitative statements about the true magnitude of the quantity concerned, may be based on subject matter knowledge and skill, general information and intuition—but not on statistical methodology. (Compare Cochran, Mosteller, and Tukey [1953, pp. 692–693].)

3.3. Definition of the Error of a Measurement, and of the Systematic Error, Precision, and Accuracy of a Measurement Process

a. Error of a Single Measurement or Adjusted Value

The *error* of any measurement of a particular quantity is, by definition, the difference between the measurement concerned and the *true value* of the magnitude of this quantity, taken positive or negative accordingly as the measurement is greater or less than the true value. In other words, if x denotes a single measurement of a quantity, or an adjusted value derived from a specific set of individual measurements, and τ is the *true value* of the magnitude of the quantity concerned, then, by definition,

the *error* of x as a measurement of $\tau \equiv x$-τ.

The error of any particular measurement or adjusted value, x, is, therefore, a fixed number. The

numerical magnitude and sign of this number will ordinarily be unknown and unknowable, because the true value of the magnitude of the quantity concerned is ordinarily unknown and unknowable. Limits to the error of a single measurement or adjusted value may, however, be inferred from (a) the *precision*, and (b) bounds on the *systematic error*, of the measurement process by which it was produced—but not without risk of being incorrect, because, quite apart from the inexactness with which bounds are commonly placed on the systematic error of a measurement process, such limits are applicable to the error of a single measurement or adjusted value, not as a unique individual outcome, but only as a typical case of the errors characteristic of measurements of the same quantity that might have been, or might be, yielded by the same measurement process under the same conditions.

b. Systematic Error of a Measurement Process

When the limiting mean μ associated with measurement of the magnitude of a quantity by a particular measurement process does not agree with the *true value* τ of the magnitude concerned, the measurement process is said to have a *systematic error*, or *bias*, of magnitude $\mu-\tau$.

The systematic error of a measurement process will ordinarily have both constant and variable components. Consider, for example, measurement of the distance between two points by means of a graduated metal tape [Holman 1892, p. 9]. Possible causes of systematic error that immediately come to mind are:

(1) Mistakes in numbering the scale divisions of the tape;
(2) irregular spacing of the divisions of the tape;
(3) sag of tape;
(4) stretch of tape;
(5) temperature not that for which the tape was calibrated.

For *any single distance*, the effects of (1) and (2) will be constant; and the effects of (3) and (4) will undoubtedly each contain a constant component characteristic of the distance concerned. Some of these effects will be of one sign, some of the other, and their algebraic sum will determine the *constant error* of this measurement process with respect to the particular distance concerned. Furthermore, the "constant error" of this measurement process will be different (at least, conceptually) for different distances measured.

In the case of repeated measurement of a single distance, the effect of (5), and at least portions of the effects of (3) and (4), may be expected to vary from one "occasion" to the next (e.g., from day to day), thus contributing *variable components* to the *systematic error* of the process.

A large fraction of the variable contributions of (3) and (4) could, and in practice no doubt would, be removed by stretching the tape by a spring balance or other means so that it is always under the same tension. The stretch corresponding to a particular distance would then be nearly the same at all times,

and a fixed correction could be made for most of the sag corresponding to this distance. Furthermore, the effect of (5) could, and in practice probably would, be reduced by determining the temperature of the tape at various points along its length and applying a temperature correction. By comparison of the tape with a standard, the error arising from (1) could be eliminated entirely, and corrections determined as a basis for eliminating, or at least, reducing the effect of (2).

As in the foregoing example there are usually certain obvious sources of systematic error. Unfortunately, there are generally additional sources of systematic error, the detection, diagnosis, and eradication of which call for much patience and acumen on the part of the observer. The work involved in their detection, diagnosis, and eradication often far exceeds that of taking the final measurements, and is sometimes discouraging to the experienced observer as well as to the beginner. Fortunately, there are various statistical tools that are helpful in this connection, and Olmstead [1952] has found that of these the two most effective and universally useful are the average ($\bar{x}$) and range (R) charts of industrial quality control. (For details on the construction and use of $\bar{x}$- and R-charts, see, for example, the ASTM Manual on Quality Control of Materials [ASTM 1951, pp. 61–63 and p. 83]; or American Standards Z1.2–1958 and Z1.3–1958 [ASA 1958b, ASA 1958c].)

c. Concept of True Value

In the foregoing we have defined the *error* of a measurement x to be the difference $x-\tau$ between the measurement and the *true value* τ of the magnitude of the quantity concerned; and the *systematic error*, or *bias*, of a measurement process as the difference $\mu-\tau$ between the limiting mean μ associated with the measurement of a particular quantity by the measurement process concerned, and the *true value* τ of the magnitude of this quantity. This immediately raises the question: Just how is the "true value" of the magnitude of a particular property of some thing defined? In the final analysis, the "true value" of the magnitude of a quantity is defined by agreement among experts on an *exemplar method* for the measurement of its magnitude—it is the limiting mean of a conceptual *exemplar process* that is an ideal realization of the agreed-upon exemplar method. And the refinement to which one should go in specifying the exemplar process will depend on the purposes for which a determination of the magnitude of the quantity concerned is needed—not just the immediate purpose for which measurements are to be taken but also the other uses to which these measurements, or a final adjusted value derived therefrom, may possibly be put.

Consider, for example, the "true value" of the length of a particular gage block. In our minds we envisage the gage block as a rectangular parallelepiped, and its *length* is, of course, the distance between its two "end" faces. But it is practically certain that the particular gage block in question is not an exact rectangular parallelepiped; and that

its two end faces are not planes, nor even absolutely smooth surfaces. Shall we define the "true length" of this gage block to be the distance between the "tops" of the highest "mountains" at each end, i.e., the distance between the two "outermost points" at each end? If so, is this distance to be measured diagonally, if necessary, or parallel to the "length-wise axis" of the gage block? If the latter, then we have the problem of how this "length-wise axis" is to be defined, especially in the case of a thin gage block whose *length* corresponds to what would ordinarily be considered to be its thickness. Or shall we be, perhaps, more sophisticated, and envisage a "mean plane" at each end, which in general will not be parallel to each other, and define the length of this gage block to be the distance between two particular points on these planes. If we choose the "outermost points" we again have the problem of the direction in which the distance is to be measured. Alternatively, we might define the length of this gage block to be the distance between two strictly parallel and conceptually perfect optical flats "just touching" the gage block at each end. If so, then is the "true distance" between these flats defined in terms of wavelengths of light via the techniques of optical interferometry the "true length" of the gage block appropriate to the purposes for which the gage block is to be used, namely, to calibrate gages and to determine the lengths of other objects by *mechanical* comparisons? Furthermore, it is clear, that the intrinsic difficulty of defining the "true value" of the *length* of a particular gage block is not eliminated if, instead, we undertake to define the "true value" of the *difference in length* of two particular gage blocks, one of which is a standard, the *accepted value* of whose length is, say, m microinches *exactly*, by industry, national or international agreement.

Similar difficulties arise, of course, in the definition of the "true value" of the *mass* of a mass standard, one of which has been resolved by international agreement. In defining the "true value" of the *mass* of a particular metallic mass standard, shall *the* mass of this particular standard be envisaged as the mass of its metallic substance alone, relative to the International Prototype Kilogram, or as the mass of its metallic substance plus the mass of the air and water vapor adsorbed upon its surface under standard conditions? The difference amounts to about 45 μg in the case of a platinum-iridium standard kilogram, and becomes critical in the case of 500 mg standards. *The mass* of a mass standard is, therefore, specified in measurement science to be the mass of the metallic substance of the standard *plus* the mass of the average volume of air adsorbed upon its surface under standard conditions. Definition of the "true value" of the *mass* of a mass standard, and *a fortiori*, of the *difference in mass* of two mass standards is, therefore, a very complex matter.

W. Edwards Deming uses the expression "preferred procedure" for what we have termed an "exemplar method," and very sagely remarks that "a preferred procedure is distinguished by the fact that it supposedly gives or would give results nearest to what are needed for a particular end; and also by the fact that it is more expensive or more time consuming, or even impossible to carry out," adding that "as a preferred procedure is always subject to modification or obsolescence, we are forced to conclude that *neither the accuracy nor the bias of any procedure can ever be known in a logical sense.*" [Deming 1950, pp. 15–17.]

It should be evident from the foregoing that the "true value" of the magnitude of some property of a thing or system cannot be defined with complete absolute exactitude.

As Cassius J. Keyser has remarked, "Absolute certainty is a privilege of uneducated minds—and fanatics. It is, for scientific folk, an unattainable ideal." [Keyser 1922, p. 120.] The degree of refinement to which one will, or ought, to go in a particular instance will depend on the uses for which knowledge of the magnitude of the property concerned is needed. The "true value" of the length of a piece of cloth in everyday commerce is certainly a fuzzy concept. "Certainly we are not going to specify that the cloth shall be measured while suspended horizontally under a tension of x pounds, at an ambient temperature of y degrees and a relative humidity of z percent" [Simon 1946, p. 654]. On the other hand, a moderate degree of refinement is necessary in defining the "true length" and "true width" of the recessed area in a window sash to which a pane of glass is to be fitted. Considerably greater refinement is needed in the definition of the "true value" of the *length* of a gage block, of the *mass* of a mass standard or of the *frequency* of a frequency standard—and in the last mentioned case there is not today, I understand, complete agreement among experts on the matter.

Indeed, as is evident from the foregoing, the "true value" of the magnitude of a particular quantity is intimately linked to the purposes for which a value of the magnitude of this quantity is needed, and its "true value" cannot, in the final analysis, be defined meaningfully and usefully in isolation from these needs. Therefore, as this fact becomes more widely recognized in science and engineering, I hope that the traditional term "true value" will be discarded in measurement theory and practice, and replaced by some more appropriate term such as "target value" [6] that conveys the idea of being the value that one would like to obtain for the purpose in hand, without any implication that it is some sort of permanent constant preexisting and transcending any use that we may have for it. I have retained the traditional expression "true value" in the sequel because of its greater familiarity, but shall always mean by it the relevant "target value."

[6] "We admit the existence of systematic error—of a difference between the quantity measured (the measured quantity) and the quantity of interest (the target quantity). We ask the observations about the measured quantity. We ask our subject matter knowledge, intuition, and general information about the relation between the measured quantity and the target quantity." [Cochran, et al. 1954, p. 33.]

". . . . . Some people prefer the term 'true value', although others excoriate it as philosophically unsound."

"We could also call the reference level a 'target value'. In a way this is a bad term because it implies that it is something we want to find through the measurement process rather than something we ought to find because, like Mt. Everest, it is there. Unfortunately our desires can influence our notion of what is true, and we can even unconsciously bring the latter into agreement with the former; my use of the term 'target value' is not meant to imply that I think it legitimate to equate what we would like to see with what is there." [Murphy 1961, p. 265.]

d. Concepts of the Precision and Accuracy of a Measurement Process

By the *precision* of a measurement process we mean the degree of mutual agreement characteristic of independent measurements of a single quantity yielded by repeated applications of the process under specified conditions; and by its *accuracy* the degree of agreement of such measurements with the true value of the magnitude of the quantity concerned. In other words, the *accuracy* of a measurement process refers to, and is determined by the degree of conformity *to the truth* that is characteristic of independent measurements of a single quantity produced (or producible) by the repeated applications of the process under specified conditions; whereas its *precision* refers solely to, and is determined solely by the degree of conformity *to each other* characteristic of such measurements, irrespective of whether they tend to be close or far from the truth. Thus, *accuracy* has to do with *closeness to the truth; precision*, only with *closeness together*

This distinction between the meanings of the terms "accuracy" and "precision" as applied to measurement processes and measuring instruments is consistent with the etymological roots of these words. "Etymologically the term 'accurate' has a Latin origin meaning 'to take pains with' and refers to the care bestowed upon a human effort to make such effort what it *ought* to be, and 'accuracy' in common dictionary parlance implies freedom from mistakes or exact conformity to *truth*. 'Precise,' on the other hand, has its origin in a term meaning 'cutoff, brief, concise'; and 'precision' is supposed to imply the property of determinate limitations or being exactly and sharply defined." [Shewhart 1939, p. 124.] Thus one can properly speak of a national, state, or local law as being "precise," but not as being "accurate"—to what *truth* can it conform? On the other hand, if one spoke of a particular translation as being "accurate" this would imply a high degree of fidelity to the original "attained by the exercise of care." Whereas, to speak of it as being "precise," would imply merely that it is unambiguous, without indicating whether it is or is not correct.[7]

In spite of the distinct difference between the etymological meanings of the terms "accuracy" and "precision," they are treated as synonyms in many standard dictionaries; and Merriam-Webster [1942], after drawing the helpful distinctions quoted in the foregoing footnote, promptly topples the structure so carefully built by adding "scrupulous exactness" as an alternative meaning of "precise." Consequently it is not surprising that "There are probably few words as loosely used by scientists as *precision* and *accuracy*.—It is not unusual to find them used interchangeably in scientific writings." [Schrock 1950, p. 10.]

[7] It is sometimes helpful to distinguish between "correct," "accurate," and "exact": "CORRECT, the most colorless term, implies scarcely more than freedom from fault or error, as judged by some (usually) conventional or acknowledged standard; . . . ACCURATE implies, more positively, fidelity to fact or truth attained by the exercise of care; . . . EXACT emphasizes the strictness or rigor of the agreement, which neither exceeds nor falls short of the fact, standard or truth; . . . PRECISE stresses rather sharpness of definition or delimitation . . ." [Merriam-Webster 1942 p. 203].

On the other hand, as Shewhart has remarked:

"Careful writers in the theory of errors, of course, have always insisted that accuracy involves in some way or other the difference between what is observed and what is true, whereas precision involves the concept of reproducibility of what is observed. Thus Laws, writing on electrical measurements, says: [8] 'Every experimenter must form his own estimate of the accuracy, or approach to the absolute truth obtained by the use of his instruments and processes of measurement. He must remember that a high precision, or agreement of the results among themselves, is no indication that the quantity under measurement has been accurately determined.' As another example we may take the following comment from a recent and authoritative treatise on chemical analysis: [9] 'The analyst should form the habit of estimating the probable accuracy of his work. It is a common mistake to confuse accuracy and precision. Accuracy is a measure of the degree of correctness. Precision is a measure of reproducibility in the hands of a given operator.' " [Shewhart 1939, pp. 124–125.]

More recently, Lundell, Hoffman, and their associates at the National Bureau of Standards have reemphasized the importance of the distinction between "precision" and "accuracy":

"In discussions of chemical analysis, the terms precision and accuracy are often used interchangeably and therefore incorrectly, for precision is a measure of reproducibility, whereas accuracy is a measure of correctness. The analyst is vitally interested in both, for his results must be sufficiently accurate for the purpose in mind, and he cannot achieve accuracy without precision, especially since his reported result is often based on one determination and rarely on more than three determinations. The recipient of the analysis is interested in accuracy alone, and only in accuracy sufficient for his purposes." [Hillebrand et al., 1953, p. 3.]

It is most unfortunate that in everyday parlance we often speak of "accuracy and precision," because *accuracy* requires *precision*, but *precision* does not necessarily imply *accuracy*.

"It is, in fact, interesting to compare the measurement situation with that of a marksman aiming at a target. We would call him a precise marksman if, in firing a sequence of rounds, he were able to place all his shots in a rather small circle on the target. Any other rifleman unable to group his shots in such a small circle would naturally be regarded as less precise. Most people would accept this characterization whether either rifleman hits the bull's-eye or not. "Surely all would agree that if our man hits or nearly hits the bull's-eye on all occasions, he should be called an accurate marksman. Unhappily, he may be a very precise marksman, but if his rifle is out of adjustment, perhaps the small circle of shots is centered at a point some distance from the bull's-eye. In that case we might regard him as an inaccurate marksman. Perhaps we should say that he is a potentially accurate marksman firing with a faulty rifle, but speaking categorically, we should have to say that the results were inaccurate." [Murphy 1961, p. 265.]

It follows from what has been said thus far that "if the precisions of two processes are the same but the biases are different, the process of smaller bias may be said to have higher accuracy while if the biases are both negligible, the process of higher precision may be said to have higher accuracy." Unfortunately, "in other cases such a simple comparison may be impossible." [ASTM 1961, p. 1760.]

[8] Frank A. Laws, Electrical Measurements, p. 593 (McGraw-Hill, New York, N.Y., 1917).
[9] G. E. F. Lundell and J. I. Hoffman, Outlines of Methods of Chemical Analysis, p. 220 (John Wiley and Sons, New York, N.Y., 1938).

To fully appreciate the preceding statement—and especially the difficulty of comparing accuracies in some cases—let us consider figures 1 and 2, in which the origins of the scales correspond to the true value of τ of the quantity measured, so that the curves shown may be regarded as depicting the distributions of *errors* of the measurements yielded by a selection of different measurement processes. Consider first the three symmetrical distributions in the top half of figure 1. All three of these distributions are centered on zero, so that these measurement processes have no *bias*. It is evident that the process of highest precision, c, is also the process of highest accuracy; and that the process of least precision, a, is also the process of least accuracy. Since curve b in the upper half of figure 1 and curve d in the lower half have identical size and shape, the corresponding processes have the same *precision*; but process b is without bias, whereas process d has a positive bias of two units, so that process b is clearly the more *accurate*. (In particular we may note that whereas it is practically certain that process b will not yield a measurement deviating

from the truth by more than two units, exactly one-half of the measurements yielded by process d will deviate from the truth by this much or more.) Similar remarks clearly apply to processes c and e corresponding to curve c in the upper half and curve e in the lower half of figure 1, but in this instance the superiority of process c relative to process e with respect to *accuracy* is even more marked. (In particular, we may note that whereas it is practically certain that no measurement yielded by process c will deviate from the truth by as much as one unit, it is practically certain that every measurement yielded by process e will deviate from the truth by more than one unit.)

Figure 2, which is essentially the same as one given by General Simon [1946, fig. 1], portrays three measurement processes A, B, and C, differing from each other with respect to both precision and bias. Comparison of these three processes with respect to *accuracy* is not quite so simple. First, it is evident that, although process A has greater precision than process B, process B is the more accurate of the two. (In particular, it is practically certain that none of the measurements yielded by process B will deviate from the truth by more than 4 units, whereas 50 percent of the measurements from process A will deviate from the truth by four units or more.) Next, is process B more (or less) *accurate* than process C which is *unbiased*, but has a very low *precision*? Process B has a positive *bias* of two units, but has sufficiently greater *precision* than process C to also have greater *accuracy* than process C. (While approximately 50 percent of the measurements
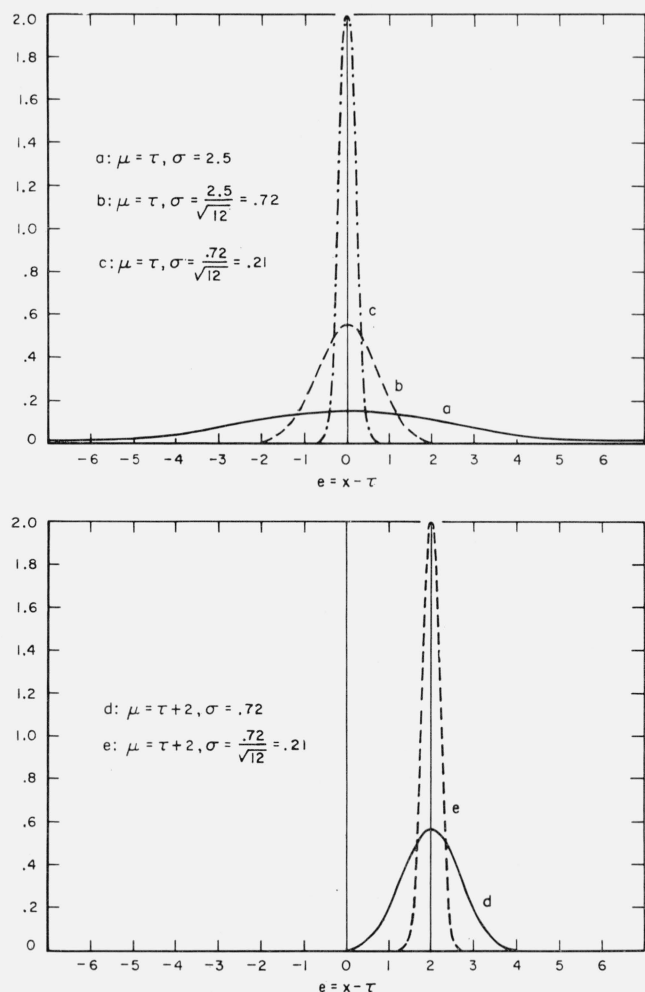

FIGURE 1. *Distributions of errors of some biased and unbiased measurement processes of various precisions.*
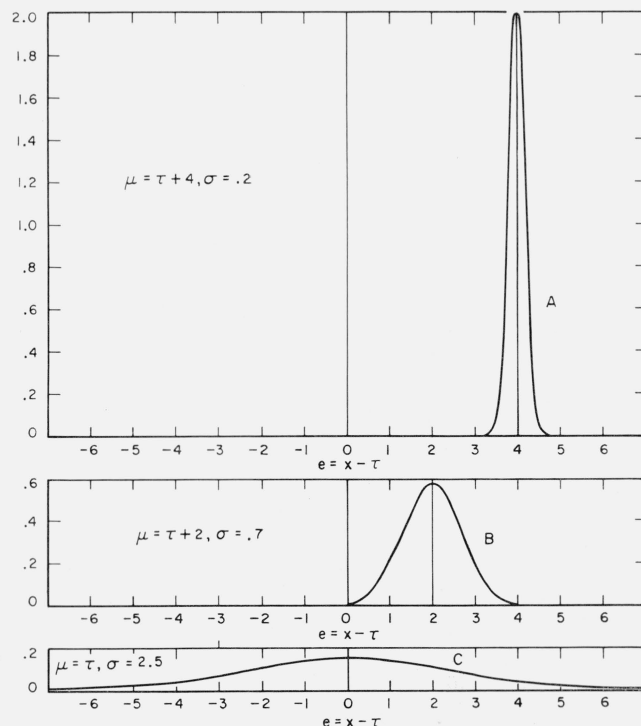

FIGURE 2. *Three measurement processes differing from each other with respect to both precision and accuracy.*

yielded by process C will deviate from the truth by more than two units (in either direction), and exactly 50 percent of the measurements yielded by process B will deviate from the truth by two units or more (in the positive direction only), it cannot be ignored that about 10 percent of the measurements yielded by process C will deviate from the truth by four units or more whereas it is practically certain that no measurement yielded by process B will deviate from the truth by as much as four units.) Similarly, it may be argued that process A, in spite of its bias, has greater *accuracy* than process C "since the range in measurements of C more than covers the corresponding ranges of A or B." [Simon 1946, p. 654.] While this conclusion that of the three measurement processes depicted in figures 2, process C has the least *accuracy*, may not be entirely acceptable to some persons, it is consistent with Gauss' dictum, in a letter to F. W. Bessel, to the effect that maximizing the probability of a zero error is less important than minimizing the "average" injurious effects of errors in general. [C. F. Gauss, 1839, pp. 146–147.]

Before leaving figure 2, we must not fail to join General Simon in remarking that "the average of a large number of measurements from [process] C will be more accurate than a similar average from either A or B" [Simon 1946, p. 654]. This point is actually illustrated in our figure 1: the three curves in the top half of figure 1 portray the distributions of errors of *single* measurements (curve *a*) of *averages of 12* measurements (curve *b*) and *averages of 144* measurements (curve *c*) from process C; and curves *d* and *e* in the lower half show the distributions of errors of *individual* measurements (curve *d*), and of *averages of 12* measurements (curve *e*) from process B, respectively. It is evident that *averages of 12* measurements from process C (curve *b* in upper portion of fig. 1) have not only greater *accuracy* than *individual* measurements from process B (curve *d* in lower portion of the figure), but also greater *accuracy* than *averages of 12* measurements from process B (curve *e* in lower portion).

On the other hand, it is obvious that, if our choice is between individual measurements from process C (curve *a*) and *averages of 12* measurements from process B (curve *e*), the latter will clearly provide greater *accuracy*. In brief, *a procedure with a small bias and a high precision can be more accurate than an unbiased procedure of low precision.* It is important to realize this, for in practical life it is often far better to always be quite close to the true value than to deviate all over the place in individual cases but strictly correct "on the average," like the duck hunter who put one swarm of shot ahead of the duck, and one swarm behind, lost his quarry, but had the dubious satisfaction of knowing that in theory he had hit it "on the average." This we must remember: in practical life we rarely make a very large number of measurements of a given type—we can't wait to be right on the average—our measurements must stand up in individual cases as often as possible.

Despite the foregoing, freedom from bias, that is, freedom from "large" bias, is a desirable characteristic of a measurement process. After all we want our measurements to yield us a determination that we can use as a substitute for the unknown value of a particular magnitude whose value we need for some purpose—we don't want a determination of the value of some other magnitude whose relation to the one we need is indefinitely known.

In view of the difficulty of comparing with respect to *accuracy* measurement processes that differ both in *bias* and *precision*, some writers have elected to take the easy way out by defining "accuracy" to be equivalent to absence of bias, saying that of two measurement processes having different biases, the process of smaller bias is the more "accurate" regardless of the relation of their respective precisions. (See, for example, Beers [1953, p. 4], Ostle [1954, p. 4], and Schenck [1961, p. 4, p. 14].) While the adoption of this concept of "accuracy" certainly makes the discussion of "accuracy" and "precision" simpler for the authors concerned, this practice is contrary to the principle of "conservation of linguistic resources," as R. B. Murphy puts it, adding: "It seems to me that the terms 'bias' and 'systematic error' are adequate to cover the situation with which they are concerned. If, nevertheless, we add the term 'accuracy' to apply again in this restricted sense, we are left wordless—at the moment at least—when it comes to the idea of over-all error. From the point of view of the need for a term it is hard to defend the view that accuracy should concern itself solely with bias. . . . [and] there is overwhelming evidence that we need a term at least for the concept of over-all error." [Murphy 1961, pp. 265–266.]

3.4. Mathematical Specification of the Precision of a Measurement Process

a. Simple Statistical Control

Let us now consider the mathematical definition of the *precision* of a measurement process under a fixed set of circumstances. By definition, the precision of a measurement process has to do with the "closeness together" that is typical of successive measurements of a single quantity generated by applications of the process under these fixed conditions. Otherwise expressed, it has to do with the typical "closeness together" of the two individual measurements constituting an arbitrary pair. If the expression *"typical 'closeness together' "* is to be meaningful, the measurements generated by repeated application of the process to the measurement of a single quantity must be homogeneous in some sense. Therefore, for the moment, let us assume that the measurement process is in a state of *simple statistical control*, so that the successive measuremets in each of the sequences (1), ($i=1, 2, 3, \ldots$), generated by the process may *all* be regarded as "observed" values of independent identically distributed random variables.

Just as we may regard each individual measurement x_{ij} in a particular sequence (1) as striving to express the value of the limiting mean μ, so also we may regard each individual difference $x_{ij}-x_{ik}, j \neq k$, as striving to express the characteristic spread between an arbitrary pair of measurements, x' and

x'', say. For this purpose the signs of these differences are clearly irrelevant. Therefore, by analogy with our use of a sequence of cumulative arithmetic means, (2), to achieve a mathematical formulation of the concept of a limiting mean associated with measurement of a given quantity by a particular measurement process, let us adopt the sequence of cumulative arithmetic means of the *squares* of the $n(n-1)/2$ distinct differences among the first n measurements of a particular sequence (1), for example, the sequence

$$(\overline{d^2})_{in} \equiv \frac{2}{n(n-1)} \sum_{j=1}^{n-1} \sum_{k=j+1}^{n} (x_{ij}-x_{ik})^2,$$
$$(n=2,3\ldots), \quad (3)$$

as the basis of a mathematical formulation of the concept of the precision of a measurement process.

The necessary and sufficient condition for almost sure convergence of the sequence (3) to a finite limit, say Δ^2, is that the Strong Law of Large Numbers be applicable to the sequence.

$$x_{i1}^2, x_{i2}^2, \ldots, x_{ij}^2 \ldots, \quad (4)$$

consisting of the squares of the corresponding terms of the original sequence (1). (Boundedness of the x's in addition to statistical control is, for example, sufficient to ensure that the sequence (4) will also obey the Strong Law of Large Numbers.) If the Strong Law of Large Numbers is applicable to the sequence of squares (4), and if the measurement process is in a state of simple statistical control, then the cumulative arithmetic means of the squares of the measurements, that is, the sequence

$$(\overline{x^2})_{in} \equiv \sum_{j=1}^{n} x_{ij}^2/n, \qquad (n=1,2,\ldots), \quad (5)$$

will almost surely tend to a limit, say S, the magnitude of which will depend on the quantity measured, the measurement process involved, but not on the "occasion" (identified by the subscript "i"). By virtue of an algebraic identity that is well known to students of mathematical inequalities, namely,

$$n\sum_{j=1}^{n} a_j^2 - \left(\sum_{j=1}^{n} a_j\right)^2 = \frac{1}{2}\sum_{j=1}^{n}\sum_{k=1}^{n}(a_j-a_k)^2, \quad (n\geq 2) \quad (6)$$

and of the fact that the right-hand side of (6) is always positive except when the a's are all equal, it is easily seen, on dividing both sides of (6) by n^2, that S will always exceed μ^2, the square of the (almost sure) limit of the sequence (2), so that we may write $S=\mu^2+\sigma^2$, with $\sigma^2>0$. Furthermore, applying the algebraic identity (6) in reverse to the right-hand side of (3) yields the following relationship between the corresponding terms of sequences (3), (5), and (1):

$$(\overline{d^2})_{in}=2\left(\frac{n}{n-1}\right)\left\{(\overline{x^2})_{in}-(\bar{x}_{in})^2\right\}>0, \quad (n\geq 2). \quad (7)$$

Hence, if a measurement process is in a state of simple statistical control and the Strong Law of Large Numbers is applicable to a *sequence* of squared measurements (4), then the sequence $(\overline{d^2})_{in}$, defined by (3), will, in view of (7), tend almost surely to a finite limit $\Delta^2=2\sigma^2$. Thus we see that σ^2, termed the *variance* of the measurement process, is the mean value of one-half of the squared difference between two arbitrary measurements x' and x'', that is,

$$\sigma^2=\frac{1}{2}\overline{(x'-x'')^2}, \quad (8)$$

and provides an indication of the *imprecision* of the process. The square root of the variance, σ, is termed the *standard deviation* of the process.

It is natural, therefore, on the basis of a single sequence of n measurements of a single quantity, to take

$$s^2 \equiv \frac{1}{2}\overline{(d^2)}=\frac{1}{n(n-1)}\sum_{j=1}^{n-1}\sum_{k=j+1}^{n}(x_j-x_k)^2=\frac{\sum_{j=1}^{n}(x_j-\bar{x})^2}{n-1}$$
$$(9)$$

as the sample estimate of the underlying variance σ^2; and the square root, s, as the sample estimate of σ.[10]

From (9), since $\bar{x}\equiv\bar{x}_n$ tends (almost surely) to μ it is evident that σ^2 is also the mean value of the squared deviations of individual measurements from the limiting mean μ of the process, that is $\sigma^2=\overline{(x-\mu)^2}$, so that the standard deviation σ may be regarded, in the language of mechanics, as the radius of gyration of the distribution of all possible measurements x about μ, the limiting mean of the process.

Remark: Mathematically the foregoing discussion can be carried out equally well in terms of the absolute (unsigned) values of the differences instead of in terms of their squares. Such an approach is, mathematically speaking, somewhat more general in that it requires for its validity merely that the Strong Law of Large Numbers be applicable to the sequence $|x_{i1}|, |x_{i2}|, \ldots, |x_{ij}|, \ldots$ of *absolute values* of the x_{ij} rather than to the sequence (4) of their squares. From the practical viewpoint, however, this greater generality is entirely illusory, and the mathematics of absolute values of variables is always more cumbersome than the mathematics of their squares. For example, the arithmetic mean of the absolute values of the $n(n-1)/2$ *distinct* differences among n measurements, i.e.,

$$\overline{|d|}_n \equiv \frac{2}{n(n-1)}\sum_{j=1}^{n-1}\sum_{k=j+1}^{n}|x_j-x_k| \quad (10)$$

---

[10] From the algebraic identity (6), it is evident that the practice in some circles of dividing $\sum_{j=1}^{n}(x-\bar{x})^2$ by n, instead of $n-1$, amounts to including each of the *distinct* squared differences $(x_j-x_k)^2$, $j\neq k$, *twice* in the summation, together with n identically zero terms $(x_j-x_k)^2$, $j=k$, each included once, and then dividing by n^2, the total number of terms (real and phantom) involved. Viewed in this light it would seem that division by $n-1$ is more reasonable, in that the inclusion of identically zero terms in the formulation of a measure of *variation* is a bit unreasonable.

is *not* expressible as a multiple of the sum of the absolute deviations of the measurements from their mean, $\sum |x_i - \bar{x}|$, and for large values of n the evaluation of (10) presents computational difficulties. The approach in terms of the absolute values of the differences also has the disadvantage from the practical viewpoint that, as we shall see in a moment, *components of imprecision* are additive in terms of squared quantities such as σ^2, so that in this sense the *variance* σ^2 is a more appropriate measure of the dispersion of the x's about their limiting mean μ than is σ itself.

Ordinarily, the magnitude of σ^2 (and, hence, of σ), unlike that of μ, depends only on the measurement process concerned and the circumstances under which it is applied, and not also on the magnitude of the quantity measured—otherwise we could not speak of a measurement process having *a* variance, or *a* standard deviation.

Since the *precision* of the process obviously decreases as the value of σ (or, of σ^2) increases, and vice versa, it is necessary to take some inverse function of σ as a measure of the precision of process. To conform with traditional usage it is necessary to regard the precision of a measurement process as *inversely proportional* to its *standard deviation* σ which is, therefore, a measure of the *imprecision* of the process. Thus, Gauss, writing in 1809, remarked that his constant $h = 1/\sigma\sqrt{2}$ could properly be considered to be a measure of the precision of the observations because if, for example $h' = 2h$, that is, if $\sigma' = \frac{1}{2}\sigma$, then "a double error can be committed in the former system with the same facility as a single error in the latter, in which case, according to the common way of speaking, a double degree of precision is attributed to the latter observations." [11]

The fact of the matter is, however, that:

". . . different fields have particularly favorite ways of expressing precision. Most of these measures are multiples of the standard deviation; it is not always clear which multiple is meant. . . .
"Some consider it unfortunate that precision should be stated as a multiple of standard deviation, since precision should increase as standard deviation decreases. Indeed, it would be more exact to say that standard deviation is a measure of imprecision. However, sensitivity, as we have previously indicated, suffers from this logical inversion without hurt. Perhaps we can best avoid this by saying that standard deviation is an index of precision. The habit of saying 'The precision is . . .' is deeply rooted, and there would be understandable impatience with the notion that standard deviation should be numerically inverted before being quoted in a statement of precision." [Murphy 1961, pp. 266–267.]

In consequence the ASTM has, at least tentatively, taken the following position:

"The numerical value of any commonly used index of precision will be smaller the more closely bunched are the individual measurements of a process. As more causes are added to the system, the greater the numerical value of the index of precision will ordinarily become. If the same index of precision is used on two different processes based on the same method or intended to measure the same physical property, the process that has the smaller value of the index of precision is said to have higher precision. Thus, although the more usual indexes of precision are really direct measures of *im*precision, this inversion of reference has been firmly established by custom. The value of the selected index of precision of a process is referred to simply as its precision or its stated precision." [ASTM 1961, p. 1759.]

As we have remarked previously, in practical work the end result of measuring some quantity or calibrating an instrument for a standard rarely consists of a single measurement of the quantity of interest. More often it is some kind of average or adjusted value, for example, the arithmetic mean of a number of independent measurements of the quantity of interest. Let us, therefore, consider the statistical properties of a sequence of *arithmetic means* of successive nonoverlapping groups of n measurements each from a sequence (1) of *individual* measurements yielded by a measurement process on a particular occasion. In other words, let us consider the sequence

$$\bar{x}_{i1}, \bar{x}_{i2}, \ldots, \bar{x}_{im}, \ldots \qquad (11)$$

of distinct arithmetic means of n measurements each

$$\bar{x}_{im} = \frac{1}{n} \sum_{j=(m-1)n+1}^{mn} x_{ij}, \quad (m=1, 2, \ldots), \qquad (12)$$

derived from a sequence (1) of individual measurements of a single quantity produced, or at least conceptually producible, by the measurement process concerned on, say, the ith occasion. If the "underlying measurement process" giving rise to the individual measurements x_{ij} is in a state of simple statistical control, then the "extended measurement process" giving rise to the averages $\bar{x}_{im}$ will also be in a state of simple statistical control. Consequently, the mathematical analysis of section 3.2, but with the averages $\bar{x}_{im}$ in place of the individual measurements x_{ij}, will carry through without other change. Let $\mu_{\bar{x}}$ denote the limiting mean thus associated with the "extended measurement process" giving rise to the averages $\bar{x}_{im}$ as its "individual" measurements. Since the cumulative arithmetic mean of the first m terms of the sequence (11) is the same as the cumulative arithmetic mean of the first mn terms of the sequence (1) of individual measurements, it is clear that the limiting mean $\mu_{\bar{x}}$ associated with the sequence of averages (11) is the same as the limiting mean associated with the original sequence (1) of individual measurements, that is,

$$\mu_{\bar{x}} = \mu_x = \mu. \qquad (13)$$

Similarly, the mathematical analysis at the beginning of the present section, but with the individual measurements x_{ij} in (3) thru (9), replaced by the averages $\bar{x}_{im}$, carries through essentially as before. Let $\sigma_{\bar{x}}^2$ denote the variance thus associated with the "extended measurement process" giving rise to the sequence of averages (11). As in the case of the variance σ^2 of individual measurements,

---

[11] "Ceterum constans h tamquam mensura praecisionis observationum considerari poterit. . . . Quodsi igitur e.g., $h'=2h$, aeque facile in systemate priori error duplex committi poterit, ac simplex in posteriori, in quo casu observationi ibus posterioribus secundum vulgarem loquendi morem praecisio duplex tribuitur." [Gauss 1809, Art. 178; 1871, p. 233; English translation, 1857, pp. 259–260.]

so also may $\sigma_{\bar{x}}^2$ be interpreted as the overall mean value of the squared deviation of "individual" averages $\bar{x}$ from the limiting mean $\mu_{\bar{x}}$ of the "extended process," that is,

$$\sigma_{\bar{x}}^2 = \overline{(\bar{x} - \mu_{\bar{x}})^2} = \overline{(\bar{x} - \mu)^2} \qquad (14)$$

By virtue of the algebraic identity

$$(\bar{x} - \mu)^2 = \left[\frac{1}{n}\sum_{j=1}^{n} x_j - \mu\right]^2 = \left[\frac{1}{n}\sum_{j=1}^{n}(x_j - \mu)\right]^2$$
$$= \frac{1}{n^2}\left[\sum_{j=1}^{n}(x_j - \mu)^2 + 2\sum_{j=1}^{n-1}\sum_{k=j+1}^{n}(x_j - \mu)(x_k - \mu)\right] \qquad (15)$$

it is readily seen that

$$\sigma_{\bar{x}}^2 = \frac{\sigma_x^2}{n} = \frac{\sigma^2}{n}. \qquad (16)$$

(The mean value of a sum is always the sum of the mean values of its individual terms, so that the overall mean value of the first summation inside the brackets in the last line of (15) is simply $n\sigma_x^2$. Furthermore, in the case of independent identically distributed measurements, the overall mean value of the term involving the double summation is 0.)

Since, from (16), $\sigma_{\bar{x}} = \sigma/\sqrt{n}$, it is seen that the precision of the arithmetic mean of n *independent* measurements is proportional to $\sqrt{n}$. Hence the arithmetic mean of *4* independent measurements has *double* the precision of a single measurement; the mean of *9* independent measurements, *thrice* the precision of a single measurement; and *144* independent measurements will be required if their arithmetic mean is to have a *12-fold* increase in precision over a single measurement. (But to ask for a 12-fold increase in precision is to ask for a very considerable improvement indeed, as can be seen from a comparison of curves *a* and *c* in the top half of fig. 1.)

To serve as a reminder of the distinction between the standard deviation of an individual measurement and the standard deviation of a mean $\bar{x}$, it is customary to refer to σ as the "standard *deviation*" of a single measurement x, and to $\sigma_{\bar{x}}$ as the "standard *error*" of the (arithmetic) mean $\bar{x}$.

b. Within-Occasions Control

In the foregoing it has been assumed that the individual measurements comprising the sequences (1) corresponding to the respective "occasions," ($i = 1, 2, \ldots$), could *all* be regarded as "observed values" of independent identically distributed random variables, that is, that the measurement process concerned was in a state of simple statistical control. When such is the case then any subset of n measurements is strictly comparable to any other subset of n measurements, and any two such subsets can be combined and regarded validly as a single set of $2n$ measurements. Unfortunately, as Student's comment quoted on page 167 above clearly implies, such complete homogeneity of measurement is rarely if ever met in practice. More often the situation is as described by Sir George Biddell Airy, British Astronomer Royal 1835–1881, in (to my knowledge) the first elementary book on the theory of errors and combination of observations in the English language [Airy 1861, p. 92]:

"When successive series of observations are made, day after day, of the same measurable quantity, which is either invariable . . . or admits of being reduced by calculation to an invariable quantity . . .; and when every known instrumental correction has been applied . . .; still it will sometimes be found that the result obtained on one day differs from the result obtained on another day by a larger quantity than could have been anticipated. The idea then presents itself, that possibly there has been on some one day, or on every day, some cause, special to the day, which has produced a *Constant Error* in the measures of that day."

Sir George, however, cautions against jumping to conclusions on the basis of only a few observations:

"The existence of a daily constant error . . . ought not to be lightly assumed. When observations are made on only two or three days, and the number of observations on each day is not extremely great, the mere fact, of accordance on each day and discordance from day to day, is not sufficient to prove a constant error. [And we should interject here that under such circumstances apparent over-all accordance is not sufficient to prove the absence of daily constant errors either.] The existence of an accordance analogous to a 'round of luck' in ordinary changes is sufficiently probable. . . . More extensive experience, however, may give greater confidence to the assumption of constant errors . . . first, it ought, in general to be established that there is possibility of error, constant on one day but varying from day to day. . . ." [Airy 1861, p. 93.]

The most useful statistical tools for this purpose are the control-chart techniques of the industrial quality control engineer. If in such a situation, a series of measurements obtained by measurement of a single quantity a number of times on each of several different days or "occasions" by a particular measurement process is plotted in the form of a *control chart for individuals* [ASTM 1951, pp. 76–78, and pp. 101, 105], the individual measurements so plotted will be seen to consist of "sections" identifiable with the subsequences (1) corresponding to the respective "occasions," ($i = 1, 2, 3, \ldots$), with the measurements within sections pair-wise closer together on the average than two measurements one of which comes from one section and the other from another. Such a series of measurements is clearly "out of control." If now parallel $\bar{x}$- and R-charts are constructed from these data, based on a series of samples of equal size from *within* the respective "occasions" or "sections" *only*, i.e., excluding means $\bar{x}$ and ranges R of any samples that "straddle" two occasions, and the points on the resulting $\bar{x}$-chart are clearly "out of control," then we may infer the existence of day-by-day components of error, constant, perhaps, on one day, but varying from day to day.

If points on the R-chart constructed as described are "out of control" also, then the measurement operation concerned is in a completely unstable condition and cannot be described validly as a "measure-

ment process" at all. On the other hand, if the $\bar{x}$-chart is "out of control," but the R-chart is "in control," then we may regard the measurement process as being in a state of *within-occasions control*. ("It is usually not safe to conclude that a state of control exists unless the plotted points for at least 25 successive subgroups fall within the 3-sigma control limits. In addition, if not more than 1 out of 35 successive points, or not more than 2 out of 100, fall outside the 3-sigma control limits, a state of control may ordinarily be assumed to exist." [ASA 1958c, p. 18.]) In such a situation we postulate the existence of (at least, conceptually) different limiting means μ_i for the respective "occasions" $(i=1, 2, ..)$, and a common *within-occasions variance* σ_w^2.

An unbiased estimate of the *within-occasions standard deviation* σ_w can be obtained, if desired, from the average range $\bar{R}$ used in constructing the R-chart, by means of the formula

$$\text{unbiased estimate of } \sigma_w = \bar{R}/d_2 \qquad (17)$$

where d_2 is the factor given in the d_2 column of table B2 of [ASTM 1951, p. 115] corresponding to the sample or subgroup size n used in constructing the R-chart.

Alternatively, if desired, an unbiased estimate of σ_w^2 can be obtained directly from the measurements involved by means of the formula

$$\text{unbiased estimate of } \sigma_w^2 = s_w^2 \frac{\sum_{h=1}^{k}\sum_{j=1}^{n}(x_{hj}-\bar{x}_h)^2}{k(n-1)}, \qquad (18)$$

where x_{hj} denotes the jth measurement and $\bar{x}_h$ the arithmetic mean of the n measurements of the hth subgroup, respectively, and k is the number of subgroups involved in constructing the R-chart.

c. Complex or Multistage Control

When a measurement process is not in a state of simple statistical control that satisfies the criteria of within-occasions control, that is, when the $\bar{x}$-chart (and control chart for individuals) are clearly "out of control," but the 25 or more subgroup ranges plotted on the R-chart exhibit control, then it is usually of importance to ascertain whether the measurement process concerned is possibly in a state of *complex or multistage statistical control*. For this purpose four or more measurements from each of at least 25 different occasions will be needed. Taking one sample of n successive measurements, $(4 \leq n \leq 10)$, from the available measurements corresponding to each of, say, $k (\geq 25)$ different "occasions," evaluate the arithmetic means $\bar{x}_i$ of these samples, $(i=1, 2, ..., k)$, and *treating these averages as IN-DIVIDUAL measurements* construct a control chart for these "individuals" and parallel $\bar{x}$- and R-charts as described in [ASTM 1951, Example 22, p. 101]. If the points plotted on these three control charts exhibit control, then we "act for the present as if"

the measurement process concerned is in a state of *complex or multistage statistical control* and regard the limiting means μ_i for the respective "occasions," $(i=1, 2, ...)$ as being in a state of simple statistical control with a limiting mean μ and variance σ_b^2, termed the *between-occasions* component of variance.

If in such a situation we were to form cumulative arithmetic means such as (3) of the squares of all distinct differences between arbitrary pairs of measurements from *within* each of the respective "occasions," then such cumulative arithmetic means of squares of differences would almost surely tend to $2\sigma_w^2$ in the limit as the number of pairs included tends to infinity, where σ_w^2 is the "within-occasions variance" mentioned above in connection with "within-occasions control." If, on the other hand we were to form similar cumulative arithmetic means of the squares of differences between arbitrary pairs consisting in each instance of one measurement from each of two different sections, then such a cumulative arithmetic mean of squared differences would tend almost certainly to $2(\sigma_b^2+\sigma_b^2)$ as the number of "occasions" sampled tends to infinity, where σ_b^2 is the above mentioned "between-occasions variance," i.e., the variance of the limiting means μ_i for the respective "occasions" about their limiting mean μ.

If in utilizing measurements from a measurement process that is in such a state of complex statistical control, one forms an average $\bar{x}_N$ that is the arithmetic mean of a total of $N=kn$ measurements, composed of n measurements from each of k different "occasions," then the variance of $\bar{x}_N$ will be

$$\sigma_{\bar{x}_N}^2 \equiv \overline{(x_N-\mu)^2} = \frac{1}{k}\left(\sigma_b^2+\frac{\sigma_w^2}{n}\right) \qquad (19)$$

From (19) it is clear that, if σ_b^2 is at all sizable compared to σ_w^2, then, for fixed $N=kn$, $\bar{x}_N$ will have greater precision *as a determination of* μ when based on a large number k of different occasions, with only a small number n of measurements from each occasion. Finally, setting $k=1$, we see that the mean $\bar{x}_i$, of n measurements all taken on the same occasion *considered as a determination of the overall limiting mean* μ has an *overall variance* $\sigma_{\bar{x}}^2=\sigma_b^2+(\sigma_w^2/n)$; but *considered as a determination of* μ_i, the limiting mean for the ith occasion, its variance is only σ_w^2/n. In other words, the "standard error" of a mean such as $\bar{x}_i$ is not unique, but depends on the purpose for which it is to be used.

An unbiased estimate of the overall standard deviation $\sigma_{\bar{x}_i}$ of the arithmetic mean of n measurements taken on a single "occasion" may be obtained by the procedure of formula (17) above, if desired, using the average range $\bar{R}$ employed in constructing the R-chart corresponding to the groups of averages $\bar{x}_{in}$.

Alternatively, an unbiased estimate of the overall variance $\sigma_{\bar{x}}^2$ can be obtained directly from the means $\bar{x}_i$ used in constructing the $\bar{x}$-chart, by using the formula

$$s = \frac{2}{\bar{x}} \frac{\sum_{i=1}^{k} (\bar{x}_i - \bar{\bar{x}})^2}{k-1} \tag{20}$$

where $\bar{x}_i$ is the arithmetic mean of the n successive observations from the ith "occasion," $(i=1, 2, \ldots, k)$ and $\bar{\bar{x}}$ is the arithmetic mean of these k means.

The foregoing concept of a state of *complex or multistage statistical control* can be extended readily to more complex truly "multistage" situations involving three or more "levels" of random variation.

Finally, it is evident from the foregoing that when a measurement process is in a state of complex or multistage statistical control, then the difference between two individual measurements (or the arithmetic means of n measurements) corresponding to two different "occasions" will include the difference $\mu_i - \mu_{i'}$ between the limiting means corresponding to the two particular occasions involved. In so far as such a comparison is regarded as a unique individual case, the difference $\mu_i - \mu_{i'}$ is a fixed constant and hence a systematic error affecting this comparison. On the other hand, if the difference between these two individual measurements (or these two arithmetic means) is regarded only as a typical instance of the outcomes that might be yielded by the same measurement process on *other* pairs of occasions, then the difference $\mu_i - \mu_{i'}$ may be regarded as a random component having a zero mean and variance $2\sigma_b^2$.

It goes without saying, of course, that if a control-chart analysis of the type described above is undertaken for the purpose of ascertaining whether the process is in a state of complex control, but the points plotted on the $\bar{x}$-chart are clearly "out of control," then the measurement process concerned cannot be regarded as statistically stable from occasion to occasion, and should be used only for *comparative measurement* within-occasions. Even when such a measurement process is used solely for comparative measurement within "occasions," it needs to be shown that comparative measurements or *fixed differences* are in a state of (simple or complex) statistical control, if this measurement process is to be generally valid in any absolute sense. Thus in the case of the thermometer calibration procedure mentioned in section 2.4 above, one needs to examine the results of repeated measurement, occasion after occasion, of the difference between two standard thermometers S_1 and S_2 of proven stability in order to determine whether the process is or is not in a state of simple or complex statistical control.

3.5. Difficulty of Characterizing the Accuracy of a Measurement Process

Unfortunately, there does not exist any single comprehensive measure of the accuracy (or *inaccuracy*) of a measurement process (analogous to the standard deviation as a measure of its imprecision) that is really satisfactory. This difficulty stems from the fact that "accuracy," like "true value," seems to be a reasonably definite concept on first thought, but

as soon as one attempts to specify exactly what one means by "accuracy" in a particular situation, the concept becomes illusive; and in attempting to resolve the matter one comes face to face, sooner or later, with the question: "Accurate" for what *purpose?*

Gauss, in his second development (1821–1823) of the Method of Least Squares clearly recognized the difficulty of characterizing sharply the "accuracy" of any particular procedure:

"Quippe quaestio haec per rei naturam aliquid vagi implicat, quod limitibus circumscribi nisi per principium aliquatenus arbitrarium nequit . . . neque demonstrationibus mathematicis decidenda, sed libero tantum arbitrio remittenda." [12] [Gauss 1823, Part I, Art. 6.]

Gauss himself proposed [loc. cit.] that the *mean square error* of a procedure—that is, $\sigma^2 + (\mu - \tau)^2$, where σ is its *standard deviation;* and $\mu - \tau$, its *bias*—be used to characterize its accuracy. While *mean square error* is a useful criterion for comparing the relative accuracies of measurement processes differing widely in both precision and bias, it clearly does not "tell the whole story." For example, if one were to adopt the principle that measurement processes having the same mean square error were equally "accurate," then one would be obliged to consider the measurement processes corresponding to the three curves shown in figure 3 as being of equal

[12] I am grateful to my colleague Franz Alt for the following literal translation of these phrases:
"For this question implies, by the very nature of the matter, something vague which cannot be clearly delimited except by somewhat arbitrary principle . . . nor can it be decided by mathematical demonstrations, but must be left to mere arbitrary judgment."
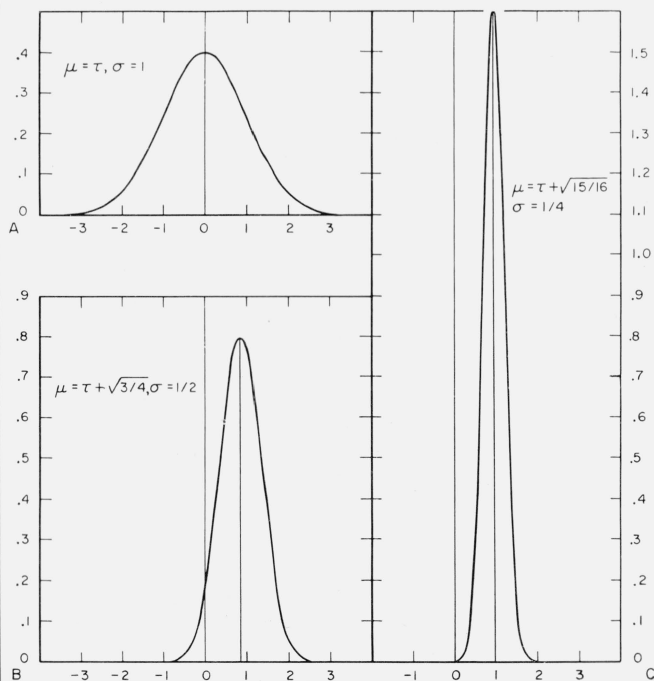

FIGURE 3. *Three distributions differing with respect to both precision and accuracy but with the same mean square error.*

accuracy, whereas for many purposes one would regard process C (portrayed to the right) as the "most accurate," in spite of the fact that the chances of scoring a "bull's eye" or "near miss" are greater in the case of process A shown in the upper left.

Alternatively, if one were to say that two measurement processes were equally accurate when exactly the same proportion P of the measurements of each lay within $\pm \delta$ units from the true value, then for $P = 0.5$ one would be obliged to say that the measurement processes corresponding to curves e and d in the lower half of figure 1 were equally accurate, and that the measurement process corresponding to curve a in the upper half of the same figure was slightly more accurate than either e or d. Or, taking $P = 0.95$, one would be obliged to say that the measurement processes corresponding to the three curves shown in figure 4 were equally accurate. From these, and other cases easily constructed, it is readily seen that it is unsatisfactory to regard two measurement processes as being equally accurate if the same specified fraction P of the measurements produced by each lie within the same distance from the true value.

Thus one is led by the force of necessity to the inescapable conclusion that ordinarily (at least) two numbers are needed to adequately characterize the accuracy of a measurement process. And this has been recognized by the American Society for Testing and Materials in their recent recommendations [ASTM 1961, pp. 1759–1760]:
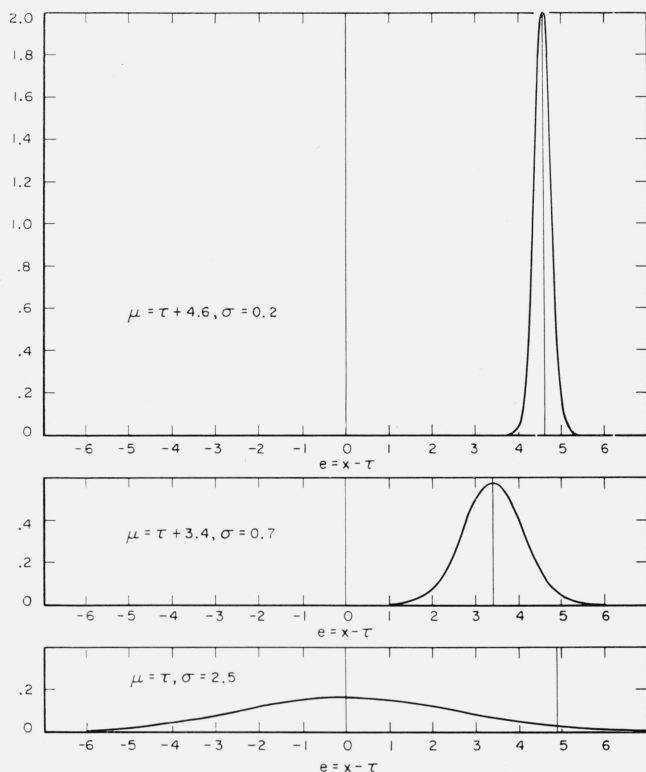
"Generally the index of accuracy will consist of two or more different numbers. Since the concept of accuracy embraces not only the concept of precision but also the idea of more or less consistent deviation from the reference level (systematic error or bias), it is preferable to describe accuracy by separate values indicating precision and bias."

The fact of the matter is that *two* numbers ordinarily suffice only because the "end results" of measurement and calibration programs are usually averages or adjusted values based on a number of independent "primary measurements," and such averages and adjusted values tend to be normally distributed to a very good approximation when four or more "primary measurements" are involved. This is illustrated by figure 5, which shows the distributions of individual measurements of two unbiased measurement processes with identical standard deviations but having *uniform* and *normal* "laws of error," respectively, together with the corresponding distributions of arithmetic means of 4 independent measurements from these respective processes— these latter two distributions are depicted by a single curve because the differences between the two distributions concerned are far less than can be resolved on a chart drawn to this scale. Since both of the processes concerned are unbiased, "accuracy" thus becomes only a matter of "precision"—or does it?—both curves for $n = 1$ have the same standard deviation, do they reflect equal "accuracy"? Would not the answer depend on the advantages to be gained from small errors balanced against the seriousness of large errors, in relation to the purpose for which a single measurement from one or the other is needed? But "the problem" disappears nicely if averages of 4 measurements are to be used.

FIGURE 4. *Three measurement processes differing in bias and precision but having 95 percent of their individual measurements within ± 4.9 units from the true value τ.*
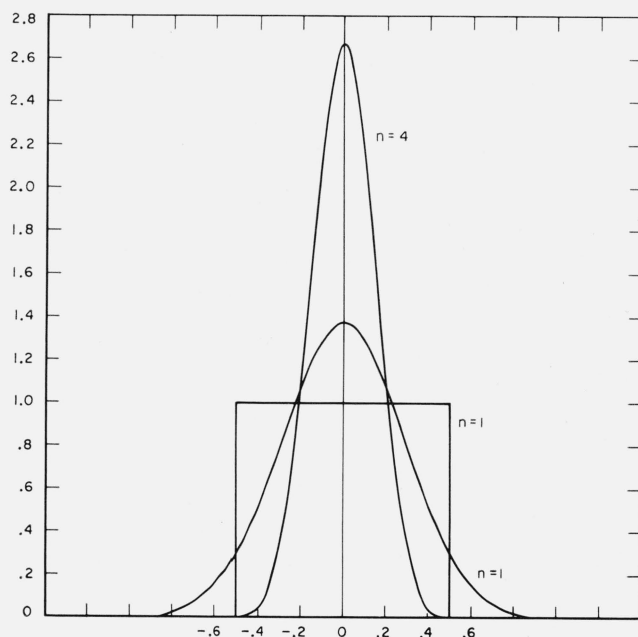

FIGURE 5. *Uniform and normal distributions of individual measurements having the same mean and standard deviation, and the corresponding distribution(s) of arithmetic means of four independent measurements.*

4. Evaluation of the Precision, and of Credible Bounds to the Systematic Error of a Measurement Process

As we have just seen, two numbers are ordinarily needed to characterize the accuracy of a measurement process, the one indicating its *precision*, and the other its *bias*. In practice, however, the bias of a measurement process is unknown and unknowable because the "true values" of quantities measured are almost always unknown and unknowable. The principle exception is when one is measuring a difference that is by hypothesis identically zero. If the bias of a measurement process could be, and were known exactly, then one would of course subtract it off as a "correction" and thus dispose of it entirely. Since ordinarily we cannot expect to know the exact magnitude of the bias of a measurement process, we are forced in practice to settle for credible bounds to its likely magnitude—much as did Steyning and the thief in chapter VI of Kipling's story, *Captains Courageous:* "Steyning tuk him for the reason that the thief tuk the hot stove—bekaze for there was nothing else that season". Consequently, neither the *bias* nor the *accuracy* of any measurement process, or method of measurement, can ever be known in a logical sense. The *precision* of a measurement process, however, can be measured and known. (Compare Deming [1950, p. 17].)

4.1. Evaluation of the Precision of a Measurement Process

In the foregoing we have stressed that a measurement operation to qualify as a *measurement process* must have attained a state of statistical control; and that until a measurement operation has been "debugged" to the extent that it has attained a state of statistical control, it cannot be regarded in any logical sense as measuring anything at all. It is also clear, from our discussion of the control-chart techniques for determining whether in any given instance one is entitled to "act for the present as if" a state of statistical control has been attained, that a fairly large amount of experience with a particular measurement process is needed before one can resolve the question in the affirmative. Once a measurement process has attained a state of statistical control, and so long as it remains in this state, then an estimate of the *standard deviation* of the process can be obtained from the data employed in establishing control, as we have indicated above.

Since the precision of a measurement process refers to, and is determined by the characteristic "closeness together" of successive independent measurements of a single magnitude generated by repeated application of the process under specified conditions, it is clearly necessary in determining whether a measurement operation is or is not in a state of statistical control, and in evaluating its precision to be reasonably definite on what variations of procedure, apparatus, environmental conditions, observers, operators, etc., are allowable in "repeated appli-

cations" of what will be considered to be the same measurement process applied to the measurement of the same quantity under the same conditions. If whatever measure of the precision and bounds to the bias of the measurement process we may adopt are to provide a realistic indication of the accuracy of this process in practice, then the "allowable variations" must be of sufficient scope to bracket the range of circumstances commonly met in practice. Scientists and engineers commonly append "probable errors" or "standard errors" to the results of their experiments and tests. These measures of imprecision are supposed to indicate the extent of the reproducibility of these experiments or tests under "essentially the same conditions," but there are great doubts whether the "probable errors" and "standard errors" generally presented actually have this meaning. The fault in most cases is not with the statistical formulas and procedures used to compute such probable errors or standard errors from the measurements in hand, but rather with the limited scope of the "conditions" sampled in taking the measurements.

a. Concept of a "Repetition" of a Measurement

As a very minimum, a "repetition" of a measurement by the same *measurement process* should "leave the door open" to, and in no way inhibit changes of the sort that would occur if, on termination of a given series of measurements, the data sheets were stolen and the experimenter were to repeat the series as closely as possible with the same apparatus and auxiliary equipment following the same instructions. In contrast, a "repetition" by the same *method of measurement* should permit and in no way inhibit the natural occurrence of such changes as will occur if the experimenter were to mail to a friend complete details of the apparatus, auxiliary equipment, and experimental procedure employed—i.e., the written text specification that defines the "method of measurement" concerned—and the friend, using apparatus and auxiliary equipment of the same kind, and following the procedural instuctions received to the best of his ability, were then, after a little practice, to attempt a repetition of the measurement of the same quantity. Such are the extremes, but there is a "gray region" between in which there is not to be found a sharp line of demarcation between the "areas" corresponding to "repetition" by the *same measurement process*, and and to "repetition" by the *same method of measurement*.

Let us consider "repetitions" by the same *measurement process* more fully. Such repetitions will undoubtedly be carried out in the same place, i.e., in the same laboratory, because if it is to be the same measurement process, the very same apparatus must be used. But a "repetition" cannot be carried out at the same *time*. How great a lapse of time should be allowed, nay *required*, between "repetitions"? This is a crucial question. Student gives an answer in a passage from which we quoted above [Student 1917, p. 415]:

"Perhaps I may be permitted to restate my opinion as to the best way of judging the accuracy of physical or chemical determinations.

"After considerable experience I have not encountered any determination which is not influenced by the date on which it is made; from this it follows that a number of determinations of the same thing made on the same day are likely to lie more closely together than if the repetitions had been made on different days.

"It also follows that if the probable error is calculated from a number of observations made close together in point of time, much of the secular error will be left out and for general use the probable error will be too small.

"Where then the materials are sufficiently stable it is well to run a number of determinations on the same material through any series of routine determinations which have to be made, spreading them over the whole period."

Another important question is: Are "repetitions" by the same measurement process, to be limited to repetitions by the same observers and operators, using the same auxiliary equipment (bottles of reagents, etc.); or enlarged to include repetitions with nominally equivalent auxiliary equipment, by various but equivalently trained observers and operators? I believe that everyone will agree that substitution, and certainly replacement, of bottles of reagents, of batteries as sources of electrical energy, etc., by "nominally equivalent materials" must be allowed. And any calibration laboratory having a large amount of "business" will certainly, in the long run at any rate, have to face up to allowing changes, even replacement of observers and operators—and, ultimately, even of apparatus.

A very crucial question, not always faced squarely, is: in complete "repetitions" by the same measurement process, are such "repetitions" to be limited to those intervals of time over which the apparatus is used "as is" and "undisturbed," or extended to include the additional variations that almost always manifest themselves when the apparatus is disassembled, cleaned, reassembled, and readjusted? Unless such disassembly, cleaning, reassembly, and readjustment of apparatus is permitted among the allowable variations affecting a "repetition" by the same measurement process, then there is very little hope of achieving satisfactory agreement between two or more measurement processes in the same laboratory that differ only in their identification with different pieces of apparatus of the same kind. In practice it is found that statistical control can be attained and maintained under such a broad concept of "repetition" only through the use of reference standards of proven stability. Furthermore, by thus more squarely facing the issue of the scope of variations allowable with respect to "repetitions" by the same measurement process, we shall go a long way toward narrowing the gap between a "repetition" by the same measurement process and by the same method of measurement.

As we have said before, if whatever measures of the precision and bias of a measurement process we may adopt are to provide a realistic indication of the accuracy of this process in practice, then the "allowable variations" must be of sufficient scope to bracket the range of circumstances commonly met in practice. Furthermore, any experimental program that aims to determine the precision and systematic error, and thence the accuracy of a measurement process, must be based on an appropriate random sampling of this "range of circumstances," if the usual tools of statistical analysis are to be strictly applicable. Or as Student put it, "the experiments must be capable of being considered to be a *random* sample of the population to which the conclusions are to be applied. Neglect of this rule has led to the estimate of the value of statistics which is expressed in the crescendo 'lies, damned lies, statistics'." [Student 1926, p. 711.]

When adequate random sampling of the appropriate "range of circumstances" is not feasible, or even possible, then it is necessary to compute, by extrapolation from available data, a more or less subjective estimate of the "precision" of the end results of a measurement operation, to serve as a substitute for a direct experimental measure of their "reproducibility." Youden [1962d] calls this "approach the 'paper way' of obtaining an estimate of the [precision]." Its validity, if any, "is based on subject-matter knowledge and skill, general information, and intuition—but not on statistical methodology" [Cochran et al. 1953, p. 693].

b. Some Examples of Realistic "Repetitions"

As Student remarked [1917, p. 415], "The best way of judging the accuracy of physical or chemical determination . . . [when] the materials are sufficiently stable . . . is . . . to run a number of determinations on the same material thru any series of routine determinations which have to be made, spreading them over the whole period." To this end, as well as to provide an overall check on procedure, on the stability of reference standards, and to guard against mistakes, it is common practice in many calibration procedures, to utilize two or more reference standards as part of the regular calibration procedure.

The calibration procedure for *liquid-in-glass thermometers*, referred to in section 2.4 above, is a case in point. A measurement of the difference between the two standards S_1 and S_2 is obtained as by-product of the calibration of the four test thermometers T_1, T_2, T_3, and T_4 in terms of the (corrected) readings of the two standards. It is such remeasurements of the difference between a pair of standard thermometers from "occasion" to "occasion" that constitutes realistic "repetitions" of the calibration procedure. The data yielded by these "repetitions" are of exactly the type needed (a) to ascertain whether or not the process is in a state of statistical control; and if so, (b) to determine its overall standard deviation.

Similarly, in the calibration of *laboratory standards of mass* at the National Bureau of Standards, "known standard weights are calibrated side-by-side with [the] unknown weights" [Almer et al., 1962, p. 33]. Indeed, weights whose values are otherwise determined "are not said to have been 'calibrated'. That term is reserved for measurements based on at least two mass standards." [loc. cit., p. 43.] In the specimen work sheets exhibited by Almer et al., the auxiliary standards involved are those from the Bureau's "NH series" of reference standards known

by the designations NH50, NH20, and NH10₁ respectively. It is the measurements obtained in routine calibrations of the differences between the values of these standards and their accepted values that not only provide valuable checks on day-to-day procedure, but also serve as the basis for determination of the overall standard deviation of this calibration process.

A third example is provided by the method followed at the National Bureau of Standards for testing *alternating-current watthour meters*, which has been described in some detail by Spinks and Zapf [1954]. Four reference watthour meters are involved. One of these, termed "the Standard Watthour Meter," is located in the device portrayed in figure 1 of the paper by Spinks and Zapf. The other three are located in a temperature-controlled cabinet. A "test" of a watthour meter sent to the Bureau involves not only a comparison of this watthour meter with the Standard Watthour Meter, but also comparisons of each of the Comparison Standard Watthour Meters with the Standard Watthour Meter. It is from the data yielded by these intercomparisons of the Standard Watthour Meter and the Comparision Standard Watthour Meters that the standard deviation of this test procedure is evaluated. Spinks and Zapf's section on "Precision and Accuracy Attainable" is notable for its exceptional lucidity as well as for its completeness with respect to relevant details.

Some additional examples of realistic "repetitions" are discussed by Youden [1962c].

4.2. Treatment of Inaccuracy Due to Systematic Errors of Assignable Origins but of Unknown Magnitudes

As we remarked in section 3.3b above, the systematic error of a measurement process will ordinarily have both constant and variable components. For convenience of exposition, it is customary to regard the individual components of the overall systematic error of a measurement or calibration process as elemental or constituent "systematic errors" and to refer to them simply as "systematic errors," for short. Included among such "systematic errors" affecting a particular measurement or calibration process are: ". . . all those errors which cannot be regarded as fortuitous, as partaking of the nature of chance. They are characteristic of the system involved in the work; they may arise from errors in theory or in standards, from imperfections in the apparatus or in the observer, from false assumptions, etc. To them, the statistical theory of error does not apply." [Dorsey 1944, p. 6; Dorsey and Eisenhart 1953, p. 104.]

The overall systematic error of a measurement process ordinarily consists of elemental "systematic errors" due to both assignable and unassignable causes. Those of unknown (not thought of, not yet identified, or as yet undiscovered) origin are always to be feared; allowances can be made only for those of recognized origin.

Since the "known" systematic errors affecting a measurement process ascribable to specific origins are ordinarily determinate in origin only, their individual values ordinarily being unknown both with respect to sign and magnitude, it is not possible to evaluate their algebraic sum and thereby arrive at a value for the overall systematic error of the measurement process concerned. In consequence, it is necessary to arrive at bounds for each of the individual components of systematic error that may be expected to yield nonnegligible contributions, and then from these bounds arrive at credible bounds to their combined effect on the measurement process concerned. Both of these steps are fraught with difficulties.

Determination of reasonable bounds to the systematic error likely to be contributed by a particular origin or assignable cause necessarily involves an element of judgment, and the limits cannot be set in exactitude. By assigning ridiculously wide limits, one could be practically certain that the actual error due to a particular cause would never lie outside of these limits. But such limits are not likely to be very helpful. The narrower the range between the assigned limits, the greater the uneasiness one feels that the assigned limits will not include whatever systematic error is contributed by the cause in question. But a decision has to be made; and on the basis of theory, other related measurements, a careful study of the situation in hand, especially its sensitivity to small changes in the factor concerned, and so forth, "the experimenter presently will feel justified in saying that he feels, or believes, or is of the opinion," that the systematic error due to the particular source in question does not exceed such and such limits, "meaning thereby, since he makes no claim to omniscience, that he has found no reason for believing" that it exceeds these limits. In other words, ' nothing has come to light in the course of the work to indicate" that the systematic error concerned lies outside the stated range. [Dorsey 1944, pp. 9–10; Dorsey and Eisenhart, 1953, pp. 105–107.]

This being done to each of the recognized potential sources of systematic error, the problem remains how to determine credible bounds to their combined effect. Before considering this problem in detail, it will be helpful to digress for a moment, to consider an instructive example relating to the combined effect of constant errors in an everyday situation.

a. An Instructive Example

Consider the hypothetical situation of an individual who is comparing his checkbook balance with his bank statement. To this end he needs to know the total value of his checks outstanding. Loathing addition, or perhaps, simply to save time, he adds up only the dollars, neglecting the cents, and thus arrives at a total of, say, $312, for 20 checks outstanding. Adding a correction of 50 cents per check, or $10 in all, he takes $322 as his estimate. Within what limits should he consider the error of this estimate to lie?

The round-off error cannot exceed ±50 cents per

check, so that barring mistakes in addition, he can be absolutely certain that the total error of his estimate does not exceed ±$10. But these are extremely pessimistic limits: they correspond to every check being in error by the maximum possible amount and all in the same direction. (Actually the maximum possible positive error is 49 cents per check or +$9.80 in all.)

To be conservative, but not so pessimistic, one might "allow" a maximum error of ±50 cents per check, but consider it reasonable to regard their signs as being equally likely to be plus or minus. In this way one would be led to conclude "with probability 0.95" that the total error lies between ±$7.00; or "with probability 0.99," between ±$8.00, as shown in the column headed "binomial" in table 1, for $n=20$. The "saving" by this procedure is clearly not great.

TABLE 1. *Limits of error of a sum of n items indicated by various methods of evaluation*

n	Absolute ±	Binomial		Uniform		Triangular		Normal, 2σ=0.5		Normal, 3σ=0.5	
		0.95 ±	0.99 ±	0.95 ±	0.99 ±	0.95 ±	0.99 ±	0.95 ±	0.99 ±	0.95 ±	0.99 ±
1	0.50	0.50	0.50	0.48	0.50	0.39	0.45	0.49	0.64	0.33	0.43
2	1.00	1.00	1.00	0.78	0.90	0.56	0.71	0.69	0.91	0.46	0.61
3	1.50	1.50	1.50	0.97	1.19	0.69	0.88	0.85	1.12	0.57	0.74
4	2.00	2.00	2.00	1.12	1.41	0.80	1.03	0.98	1.29	0.65	0.86
5	2.50	2.50	2.50	1.25	1.60	0.89	1.15	1.10	1.44	0.73	0.96
6	3.00	2.50	3.00	1.38	1.76	0.98	1.29	1.20	1.58	0.80	1.05
7	3.50	3.00	3.50	1.49	1.91	1.06	1.39	1.30	1.70	0.86	1.14
8	4.00	3.50	3.50	1.59	2.05	1.13	1.49	1.39	1.82	0.92	1.21
9	4.50	3.50	4.00	1.69	2.18	1.20	1.58	1.47	1.93	0.98	1.29
10	5.00	4.00	4.50	1.78	2.31	1.26	1.66	1.55	2.04	1.03	1.36
15	7.50	5.50	6.00	2.19	2.88	1.55	2.04	1.90	2.49	1.27	1.69
20	10.00	7.00	8.00	2.53	3.33	1.79	2.35	2.19	2.88	1.46	1.92
25	12.50	8.50	9.50	2.83	3.72	2.00	2.63	2.45	3.22	1.63	2.15
30	15.00	10.00	11.00	3.07	4.03	2.19	2.88	2.68	3.53	1.79	2.35
40	20.00	13.00	14.00	3.58	4.70	2.53	3.33	3.10	4.07	2.07	2.72
50	25.00	16.00	17.00	4.00	5.26	2.83	3.72	3.46	4.55	2.31	3.04
60	30.00	19.00	20.00	4.38	5.76	3.10	4.07	3.80	4.99	2.53	3.33

Alternatively, one might consider it to be more "realistic" to regard the individual errors as independently and uniformly distributed between −50 cents and +50 cents, concluding "with probability 0.95" that the total error does not exceed ±$2.53; or "with probability 0.99," is not greater then ±$3.33—as shown in the columns under the heading "uniform" in table 1. It is clear that a considerable reduction in the estimate of the total error is achieved by this approach.

Strictly speaking, the foregoing analyses via the theory of probability are both inapplicable to the problem at hand: each round-off error is a fixed number between ±50 cents, and their sum is a fixed number between ±$10. If it were true that round-off errors in such cases were uniformly distributed between ±50 cents, then, if one made a habit of evaluating limits of error according to this procedure, one could expect the limits of error so calculated to include the true total error in 95 percent, or 99 percent of the instances in which this procedure was used in the long run. Round-off errors in such cases are almost certainly not uniformly distributed between ±50 cents. (Many items are priced these days at $2.98 etc., and this will distort the distribution of the cents-portion of one's bills but added sales taxes no doubt have a "smoothing" effect.)

Nevertheless, I believe that you will agree that if, in the hypothetical case under discussion, the checkbook balance, with an allowance of $322 for checks outstanding, failed to agree with the bank statement to within $2.53 (or $3.33), our "friend" would do well to check into the matter more thoroughly. And, alternatively, if his checkbook balance so adjusted, and the bank statement, agreed to within $2.53 (or $3.33), it would be reasonably

"safe" for him to "act for the present as if" his balance and the bank statement were in agreement. (See Eisenhart [1947a, p. 218] for discussion of a similar example relating to computation with logarithms.)

b. Combination of Allowances for Systematic Errors

The foregoing example suggests that a similar procedure be used for arriving at credible limits to the likely overall effect of systematic errors due to a number of different origins. A number of additional difficulties confront us, however, in this case. To begin with, in view of the inexactness with which bounds can ordinarily be placed on each of the individual components of systematic error, it is not possible to say with absolute certainty that their combined effect lies between the sum of the positive bounds and the sum of the negative bounds.

Second, even if it were possible to scale the situation so that the bounds for each of the components of systematic error was the same, say, ±Δ, there would still remain the problem of translation into an appropriate probability calculus. Most persons would, I believe, regard the "binomial" approach (corresponding to equal probability of maximum error in either direction), as too pessimistic; and the approach via a uniform distribution of error, as a bit conservative, on the grounds that one intuitively feels that the individual errors are somewhat more likely to lie near the centers than near the ends of their respective ranges. Therefore, one might attempt to simulate this "feeling" by assuming the "law of error" to be an isosceles triangle centered at zero and ends at ±Δ; or, more daringly, by assuming the "law of error" to be approximately normal with Δ corresponding to 2 "σ" or even 3 "σ."

Unfortunately whatever "probability limits" may be placed upon the combined effects of several independent systematic errors by these procedures are quite sensitive to the assumption made at this stage, as is evident from table 1. Therefore, anyone who uses one of these methods for the "combination of errors" should indicate explicitly which of these (or an alternative method) he has used. When (a) the number of systematic errors to be combined is large, (b) the respective ranges are approximately equal in size, and (c) one feels "fairly sure" that the individual errors do not fall outside of their respective ranges, then my personal feeling is that the "uniform" method is probably a wee bit conservative but "safe"; the triangular method is a bit "too daring"; the normal method with "σ"=$\Delta/3$ ordinarily "much too daring"; but the normal method with "σ"=$\Delta/2$, probably "not too daring." When (b) and (c) hold but n is small, then it will probably be safe to use the "uniform" method with "Δ" taken equal to the average of the individual ranges. Other cases, e.g., when n is large but, say, one or two of the ranges is (are) much larger than the others and tend(s) to dominate the situation, requires special consideration which is beyond the scope of the present paper.

4.3. Expression of the Inaccuracy of a Measurement Process

By whatever means credible bounds to the likely overall systematic error of the measurement process are obtained they should not be combined (by simple addition, by "quadrature," or otherwise) with an experimentally determined measure of its standard deviation to obtain an overall index of its accuracy (or, more correctly, of its *inaccuracy*). Rather (a) the standard deviation of the process and (b) credible bounds to its systematic error should be stated separately, because, as we showed in figure 3, a measurement process having standard deviation σ=0.25 and a bias $\Delta=\sqrt{15/16}=0.97$ is for most purposes "more accurate" than a measurement process having zero bias and standard deviation σ=1, so that a process with σ=0.25 and a bias *less than* ± 0.97 will *a fortiori* be "more accurate."

Finally, if the uncertainties in the assigned value of a national standard or of some fundamental constant of nature (e.g., in the *volt as maintained at the National Bureau of Standards*, or in the speed of light c, or in the acceleration of gravity g on the Potsdam basis) is an important potential source of systematic error affecting the measurement process, no allowance for possible systematic error from this source should be included ordinarily in evaluating overall bounds to the systematic error of the measurement process. Since the error concerned, what ever it is, affects all results obtained by the method of measurement involved, to include an allowance for this error would be to make everybody's results appear unduly inaccurate relative to each other. Instead, in such instances one should state (a) that results obtained by the measurement process concerned are in terms of the volt (or the watthour, or the kilogram, etc.)

"as maintained at the National Bureau of Standards" [McNish and Cameron 1960, p. 102], or "correspond to the speed of light c=2.997925×10^{10} cm/sec. *exactly*," say; and (b) that the indicated bounds to the systematic error of the process are exclusive of whatever errors may be present from this (or these) source(s). Given such information, experts can make such additional allowances, as may be needed, in fundamental scientific work; and comparative measurements within science and industry within the United States will not appear to be less accurate than they very likely are for the purposes for which they are to be used.

———

5. Bibliography

Airy, George Biddell (1861), On the Algebraical and Numerical Theory of Errors of Observations and the Combination of Observations (Macmillan and Co., Cambridge and London).

Almer, H. E., L. B. Macurdy, H. S. Peiser, and E. A. Weck (1962), Weight calibration schemes for two knife-edge direct-reading balances, J. Research NBS **66C** (Eng. and Instr.) No. 1, pp. 33–44.

American Standards Association (1958a), Guide for quality control, American Standard Z 1.1–1958. (American Standards Association, 70 East Forty-fifth St., New York 17, N.Y.)

American Standards Association (1958b), Control chart method of analyzing data, American Standard Z 1.2–1958. (American Standards Association, 70 East Forty-fifth St., New York 17, N.Y.)

American Standards Association (1958c), Control chart method of controlling quality during production. American Standard Z 1.3–1958. (American Standards Association, 70 East Forty-fifth St., New York 17, N.Y.)

American Society for Testing Materials (1951), ASTM Manual on Quality Control of Materials Special Technical Publication **15**–C (American Society for Testing Materials 1916 Race St., Philadelphia 3).

American Society for Testing and Materials (1961), Use of the terms precision and accuracy as applied to measurement of a property of a material, ASTM Designation: E 177–61T. Reprinted from ASTM Standards, Pt 11, pp. 1758–1766.

Baird, D.C. (1962), Experimentation: An Introduction to Measurement Theory and Experiment Design, (Prentice-Hall, Inc., Englewood Cliffs, N.J.).

Beers, Yardley (1953), Introduction to the Theory of Error, (Addison-Wesley Publishing Co., Cambridge 42, Mass.).

Bicking, Charles A. (1952), The reliability of measured values—an illustrative example. Photogrammetric Engineering **XVIII**, pp. 554–558.

Cameron, J. M. (1951), The use of components of variance in preparing schedules for the sampling of baled wool, Biometrics **7**, pp. 83–96.

Chauvenet, William (1868), A Manual of Spherical and Practical Astronomy Vol. II, 4th edition, (J. B. Lippincott and Co., Philadelphia).

Cochran, William G., Frederick Mosteller, and John W. Tukey (1953), Statistical problems of the Kinsey report, J. Am. Stat. Assoc. **48**, pp. 673–716.

Cochran, William G., Frederick Mosteller, and John W. Tukey (1954), Principles of sampling, J. Am. Stat. Assoc. **49**, pp. 13–35.

Crow, Edwin L. (1960), An analysis of the accumulated error in a hierarchy of calibrations. IRE Trans. Instr. **I–9**, pp. 105–114.

Deming, W. Edwards and Raymond T. Birge (1937), On the Statistical Theory of Errors, reprinted from Reviews of Modern Physics **6,** pp. 119–161 (1934) with additional notes dated 1937 (The Graduate School, U.S. Department of Agriculture, Washington 25, D.C.).

Deming, W. Edwards (1943), Statistical Adjustment of Data (John Wiley & Sons, Inc., New York, N.Y.)

Deming, W. Edwards (1950), Some Theory of Sampling (John Wiley & Sons, New York, N.Y.).

Dorsey, N. Ernest (1944), The velocity of light, Transactions American Philosophical Society **XXXIV,** pp. 1–110.

Dorsey, N. Ernest and Churchill Eisenhart (1953), On absolute measurement, The Scientific Monthly **LXXVII,** pp. 103–109.

Eisenhart, Churchill (1947a), Effects of rounding or grouping data, Chapter 4 of Techniques of Statistical Analysis, edited by C. Eisenhart, M. W. Hastay, W. A. Wallis (McGraw-Hill Book Co., New York, N.Y.).

Eisenhart, Churchill (1947b), Planning and interpreting experiments for comparing two standard deviations, Chapter 8 of Techniques of Statistical Analysis, edited by C. Eisenhart, M. W. Hastay, W. A. Wallis (McGraw-Hill Book Co., New York, N.Y.).

Eisenhart, Churchill (1949), Probability center lines for standard deviation and range charts, Industrial Quality Control **vi,** pp. 24–26.

Eisenhart, Churchill (1952), The reliability of measured values—fundamental concepts, Photogrammetric Engineering **XVIII,** pp. 542–554 and 558–565.

Eisenhart, Churchill (1962), On the realistic measurement of precision and accuracy, ISA Proceedings of the Eight National Aero-Space Instrumentation Symposium held in Washington, May 1962, pp. 75–83.

Feller, William (1957), An Introduction to Probability Theory and its Applications, Vol. 1, 2d edition (John Wiley & Sons, New York, N.Y.).

Galilei, Galileo (1638), Discorsi e Dimostrazioni Matematiche Intorno a Due Nuove Scienze, Leiden.

Galilei, Galileo (1898), Discorsi e Dimostrazioni Matematiche Intorno a Due Nuove Scienze, Le Opere di Galileo Galilei (Edizione Nazionale) **VIII,** pp. 39–448, Firenze.

Galilei, Galileo (1914), Dialogues Concerning Two New Sciences, translated by Henry Crew and Alfonso de Salvio, with an Introduction by Antonio Favaro (The Macmillan Co., New York, N.Y.).

Gauss, C. F. (1809), Theoria Matus Corporum Coelestium in Sectionibus Conicis Solem Ambientium, Frid. Perthes et I. H. Besser, Hamburg; reprinted in Carl Friedrich Gauss Werke, Band **VII,** Gotha, 1871.

Gauss, C. F. (1823), Theoria Combinationis Observationum Erroribus Minimis Obnoxiae Commentationes societatis regiae scientiarum Gottingensis recentiores, **V,** pp. 1–104, Gottingae; reprinted in Carl Friedrich Gauss Werke, Band **IV,** Göttingen, 1873.

Gauss, C. F. (1839), letter to F. W. Bessel dated February 28, 1839, reproduced in "Kritische bemerkungen zur methode der kleinsten quadrate," pp. 142–148 in Carl Friedrich Gauss Werke, Band **VIII,** (B. G. Teubner, Leipzig, 1900).

Gauss, C. F. (1857), Theory of the Motion of the Heavenly Bodies Moving About the Sun in Conic Sections; English translation by Charles Henry Davis (Little, Brown and Co., Boston).

Gnedenko, B. V. (1962), The Theory of Probability (English translation by B. D. Sechler), (Chelsea Publishing Co., New York, N.Y.).

Hermach, F. L. (1961), An analysis of errors in the calibration of electric instruments, Communication and Electronics (AIEE) **54,** pp. 90–95.

Hillebrand, W. F., G. E. F. Lundell, H. A. Bright, J. I. Hoffman, Applied Inorganic Analysis, 2d ed. (1953), (John Wiley & Sons, Inc., New York, N.Y.).

Holman, Silas Whitcomb (1892), Discussion of the Precision of Measurements, (John Wiley and Sons, New York, N.Y.).

Keyser, Cassius J. (1922), Mathematical Philosophy, (E. P. Dutton and Co., New York, N.Y.).

Kline, S. J. and F. A. McClintock (1953), Describing uncertainties in single-sample experiments, Mech. Eng. **75,** pp. 3–8.

Laplace, Pierre Simon (1886), Theorie Analytique Des Probabilites; 3d edition, Vol. 7 of Oeuvres Completes de Laplace publiées sous les auspices de l'Académie des Sciences, (Gauthier-Villars, Imprimeur-Libraire de l'École Polytechnique, du Bureau des Longtudes, Successeur de Mallet-Bachelier, Quai des Grands-Augustins, 55, Paris).

McNish, A. G. and J. M. Cameron (1960), Propagation of error in a chain of standards, IRE Trans. Instr. **I–9,** pp. 101–104.

Millikan, R. A. (1903), Mechanics, Molecular Physics, and Heat, (Ginn and Co., New York, pp. 195–196).

Murphy, R. B. (1961), On the meaning of precision and accuracy, Materials Research and Standards **4,** pp. 264–267.

NPL (1957), Calibration of temperature measuring instruments, National Physical Laboratory Notes on Applied Science, No. 12, pp. 29–30, (Her Majesty's Stationery Office, London).

Ostle, Bernard (1954), Statistics in Research, (The Iowa State College Press, Ames, Iowa).

Parzen, Emanuel (1960), Modern Probability Theory and its Applications, (John Wiley & Sons, New York, N.Y.).

Proschan, Frank (1953), Confidence and tolerance intervals for the normal distribution, J. Am. Stat. Assoc. **48,** pp. 550–564.

Rossini, F. D. and W. Edwards Deming (1939), The assignment of uncertainties to the data of chemistry and physics, with specific recommendations for thermochemistry, J. Wash. Acad. Sci. **29,** pp. 416–441.

Schenck, Hilbert, Jr. (1961), Theories of Engineering Experimentation, (McGraw-Hill Book Co., Inc., New York, N.Y.).

Schrock, Edward M. (1950), Quality Control and Statistical Methods, (Reinhold Publishing Corp., New York 18, N.Y.).

Shewhart, W. A. (1931), Economic Control of Quality of Manufactured Product, (D. Van Nostrand Company, Inc., New York, N.Y.).

Shewhart, Walter A. (1939), Statistical Method from the Viewpoint of Quality Control, (The Graduate School, U.S. Department of Agriculture, Washington, D.C.).

Shewhart, Walter A. (1941), Contribution of statistics to the science of engineering, University of Pennsylvania Bicentennial Conference, Volume on Fluid Mechanics and Statistical Methods in Engineering, pp. 97–124, (University of Pennsylvania Press, Philadelphia).

Simon, Leslie E. (1941), An Engineer's Manual of Statistical Methods, (John Wiley & Sons, Inc., New York, N.Y.).

Simon, Leslie E. (1942), Application of statistical methods to ordnance engineering. J. Am. Stat. Assoc. **37,** pp. 313–324.

Simon, Leslie E. (1946), On the relation of instrumentation to quality control, Instruments **19,** pp. 654–656 (Nov. 1946); reprinted in Photogrammetric Engineering **XVIII,** pp. 566–573 (June 1952).

Spinks, A. W. and T. L. Zapf (1954), Precise comparison method of testing alternating-current watthour meters. J. Research NBS **53,** pp. 95–105.

Student (1908), The probable error of a mean, Biometrika **VI,** No. 1, pp. 1–25.

Student (1926), Mathematics and agronomy, Journal of the American Society of Agronomy **XVIII,** 703–719.

Student (1927), Errors of routine analysis. Biometrika, **XIX,** pp. 151–164.

Swindells, James F. (1959), Calibration of liquid-in-glass thermometers, NBS Circ. 600, pp. 11–12, (U.S. Government Printing Office, Washington 25, D.C.).

Tukey, J. W. (1960), Conclusions vs. decisions, Technometrics **2,** No. 4, pp. 423–433.

Waidner, C. W. and H. C. Dickinson (1907), On the standard scale of temperature in the interval 0° to 100° C, Bul. Bur. Stds. **3,** pp. 663–728.

Webster's Dictionary of Synonyms (1942, 1st ed.), (G. and C. Merriam Co., Springfield, Mass.).

Youden, W. J. (1950), Comparative tests in a single laboratory, ASTM Bulletin No. 166, pp. 48–51.

Youden, W. J. and J. M. Cameron (1950), Use of statistics to determine precision of test methods, Symposium on Application of Statistics, Special Technical Publication No. 103, pp. 27–34, (American Society for Testing Materials, Philadelphia).

Youden, W. J. (1951a), Statistical Methods for Chemists, (John Wiley & Sons, New York, N.Y.).

Youden, W. J. (1951b), Locating sources of variability in a process, Ind. Eng. Chem. **43**, pp. 2059–2062.

Youden, W. J. (1953), Sets of three measurements: The Scientific Monthly **LXXVII**, pp. 143–147.

Youden, W. J. (1954–1959), Statistical Design, Industrial and Engineering Chemistry, Feb. 1954 to Dec. 1959, Bimonthly articles collected in a single booklet available from Reprint Department, ACS Applied Publications, 1155 Sixteenth St., Washington 6, D.C. ,

Youden, W. J., W. S. Connor, and N. C. Severo (1959), Measurements made by matching with known standards, Technometrics **1**, pp. 101–109.

Youden, W. J. (1960), The sample, the procedure, and the laboratory, Anal. Chem. **32**, pp. 23A–37A.

Youden, W. J. (1961a) How to evaluate accuracy, Mat. Res. & Std. **1**, pp. 268–271.

Youden, W. J. (1961b), What is the best value? J. Wash. Acad. of Sci. **51**, pp. 95–97.

Youden, W. J. (1961c), Statistical problems arising in the establishment of physical standards, Proceedings Fourth Berkeley Symposium on Mathematical Statistics and Probability **III,** pp. 321–335 (University of California Press, Berkeley and Los Angeles.).

Youden, W. J. (1961d), Systematic errors in physical constants, Phys. Today **14,** pp. 32–42 (1961); also in Technometrics **4,** pp. 111–123 (1962).

Youden, W. J. (1962a), Experimentation and Measurement, National Science Teachers Association Vistas of Science Series No. 2, (Scholastic Book Services, New York 36, N.Y.).

Youden, W. J. (1962b), Measurement Agreement Comparisons, presented at the Standards Laboratory Conference, National Bureau of Standards, Boulder, Colo., August 8–10, 1962.

Youden, W. J. (1962c), Uncertainties in calibration, IRE Trans. Instr. **I–11,** pp. 133–138 (1962).

Youden, W. J. (1962d), Realistic estimates of errors in measurements, ISA Journal **9,** No. 10, pp. 57–58.

(Paper 67C2–128)

Publications of the National Bureau of Standards*

Selected Abstracts

Quantitative metallography with a digital computer: application to an Nb-Sn superconducting wire, G. A. Moore and L. L. Wyman, *J. Research NBS* **67A** (*Phys. and Chem.*) *No. 2, 127 (Mar.–Apr. 1963)*. **70** *cents*.

Accurate quantitative data pertinent to the structure of solid materials at the micro size level, which are difficult or prohibitive to obtain by traditional manual measurements, are now obtained directly by a digital computer which uses a photomicrograph as the information input. The history of picture interpretation experiments at the NBS is reviewed. The fundamental computer operations are illustrated, together with a description of 24 image processing routines now functional at a practical level.

A micrograph of a specimen of Nb-Sn superconductor wire is exhaustively analyzed. This specimen is found to contain approximately 70 percent Nb_3Sn, nearly all of which is mutually interconnected. It is also found that in this specimen the mean free path in the Nb_3Sn superconducting phase is only 26.5 microns. This small value results from the spongy structure of the material and numerous interruptions caused by voids and by particles of 4 other solid phases. The comparative importance of the several types of interruptions is measured. It is determined that small voids are the most important single cause of the short mean free path, and deduced that these voids appear to have formed mainly from the reaction during heat treatment.

Moiré fringes produced by a point projection x-ray microscope, S. B. Newman, *J. Research NBS* **67A** (*Phys. and Chem.*) *No. 2, 149 (Mar.–Apr. 1963)*. *70 cents*.

Moiré fringes produced by soft X-rays passing through crossed gratings of fine wire mesh are demonstrated. Regular systems of bands appearing superimposed on radiomicrographs of oriented cellulosic structures may also be moiré fringes. These fringes could be formed by fibrillate structures acting as crossed diffraction gratings.

A method for determining the elastic constants of a cubic crystal from velocity measurements in a single arbitrary direction; application to $SrTiO_3$, J. B. Wachtman, Jr., M. L. Wheat, and S. Marzullo, *J. Research NBS* **67A** (*Phys. and Chem.*) *No. 2, 205 (Mar.–Apr. 1693)*. *70 cents*.

A method is given for calculating the three elastic constants of a cubic crystal and their standard deviations from the three velocities of sound and their standard deviations measured in a single direction. The method is applicable to any direction except [100] and [111].

A new type of computable inductor, C. H. Page, *J. Research NBS* **67B** (*Math. and Math. Phys.*) *No. 1, 31 (Jan.–Mar. 1863)*. *75 cents*.

The mutual inductance analog of the generalized Thompson-Lampard theorem (for cross-capacitances) is developed. An infinitely long cage of five parallel wires can yield an absolute inductance of

$$10^{-7} \ln \frac{3+\nu 5}{2}$$

henries per meter. End-effects of order $1/l^2$ occur in a finite cage, but can be reduced to order $1/l^4$ by using eight wires.

The eight wire cage has the advantage of overdetermined relations among the inductances to be measured, allowing an estimate of experimental error in the calibration of a standard. Errors due to faulty cage geometry are shown to be of the order of 1 in 10^7.

Input admittance of linear antennas driven from a coaxia line, T. T. Wu, *J. Research NBS* **67D** (*Radio Prop.*) *No. 1' 83–89 (Jan.–Feb. 1963)*. *70 cents*.

In two cases of a linear antenna driven from a coaxial line, it is shown that the apparent terminal admittance to the coaxial line can be additively separated into two parts when the transverse dimensions are small compared with the wavelength. One of these two parts depends only on the wavelength and the dimensions of the antenna, while the other part can be interpreted as a capacitance that depends only on the radii of the coaxial line. This capacitance may be found exactly from the solution of an integral equation, in the sense that further corrections cannot be interpreted simply as a capacitance.

Corrosion of steel pilings in soils, M. Romanoff, *NBS Mono. 58 (Oct. 24, 1962), 20 cents*.

Steel pilings have been used for many years as structural members of dams, floodwalls, bulkheads, and as load-bearing foundations. While its use is presumably satisfactory, no evaluation of the material after long service has been made. In cooperation with the American Iron and Steel Institute and the U.S. Corps of Engineers, the National Bureau of Standards has undertaken a project to investigate the extent of corrosion on steel piles after many years of service.

Results of inspections made on steel pilings which have been in service in various underground structures under a wide variety of soil conditions for periods of exposure up to 40 years are presented.

In general, no appreciable corrosion of steel piling was found in undisturbed soil below the water table regardless of the soil types or soil properties encountered. Above the water table and in fill soils corrosion was found to be variable but not serious.

It is indicated that corrosion data previoulsy published by the National Bureau of Standards on specimens exposed under disturbed soil conditions do not apply to pilings which are driven in undisturbed soils.

Radiation quantities and units, *International Commission on Radiological Units and Measurements (ICRU) Report 10a, 1962, NBS Handb. 84 (Nov. 14, 1962), 20 cents*.

This Handbook presents definitions of 23 fundamental radiation quantities and units. It resulted from a 3-year study by the *Ad Hoc* Committee on Quantities and Units of the ICRU. It includes new names for certain quantities and clarified definitions for others. It presents a system of concepts and a set of definitions which is internally consistent and yet of sufficient generality to cover present requirements and such future requirements as can be foreseen.

A tabulation of the thermodynamic properties of normal hydrogen from low temperature to $300°$ K and from 1 to 100 atmospheres, J. W. Dean, *NBS Tech. Note 120 (Nov. 1961), 45 cents*.

Pressure, volume, temperature, internal energy, enthalpy, and entropy of normal hydrogen gas have been tabulated along isobars in 1 °K temperature steps. The range covered is from the saturation temperature to 300 °K and from a pressure of 1 to 100 atmospheres. The source of data is the Research Paper 1932 of the National Bureau of Standards Journal of Research. The method is described by which the data presented in Research Paper 1932 is reduced to properties directly useful for engineering calculations. A method is also described for estimating the effect of ortho-para compositions upon the tabulated properties.

Tabular values are presented in the dimensional units of the metric system. The tabulations are also available in the

dimensional units of the British system as Technical Note No. 120, Supplement A.

Controlled temperature oil baths for saturated standard cells,
P. H. Lowrie, Jr., *NBS Tech. Note 141 (Aug. 1962), 25 cents.*
Two oil baths for the temperature control of saturated standard cells have been designed and fabricated at the Boulder Laboratories of the National Bureau of Standards for operation at 28 °C and 35 °C respectively. Short term control to better than ± 0.001 °C with day-to-day variations no greater than 0.002 °C has been achieved with the use of a mercury-toluene thermoregulator incorporating a temperature anticipating device. The circulating system limits temperature gradients in the oil to less than 0.001 °C across any 10-inch section. The baths incorporate preheat and drain tanks as well as the main temperature regulated tank to facilitate the insertion and removal of cells and to minimize oil spillage.

Coordinated color identifications for industry,
K. L. Kelly, *NBS Tech. Note 152 (Nov. 1962), 15 cents.*
When a color is to be identified, the preciseness required of the identification is the first consideration. Usually this is determined by a trial-and-error method which can be both costly and time-consuming. For some uses, a color name consisting of a hue name or a hue name and modifier is sufficient while for others, a notation of the color in a color-order system will suffice. Where maximum precision is required, the color should be measured instrumentally and the results expressed numerically. This paper describes the coordinated series of five levels of fineness of color identification developed by ISCC Subcommittee for Problem 23, the Expression of Historical Color Usage, and is based on the ISCC–NBS method of designating colors. It lists the methods for changing from one level to another and gives examples of the use of each level.

The thermodynamic properties of helium from 6 to 540° R between 10 and 1,500 psia,
D. B. Mann, *NBS Tech. Note 154A (Jan. 1962), 50 cents.*
The specific volume, enthalpy, entropy, and internal energy values of helium are presented in tabular form as functions of pressure and temperature.
Data are tabulated in two-degree Rankin increments for 36 isobars between 10 psia and 1,500 psia. A comparison with previously published data is made where applicable.
An expression is presented which represents the pressure-density-temperature surface based on previously published data.
The tabulation is presented in the dimensional units of the British system but is also available in the dimensional units of the metric system.

Emission stabilization of thermionic diode noise sources,
M. W. Randall and M. G. Arthur, *NBS Tech. Note 160 (Sept. 1962), 15 cents.*
An apparatus is described which is capable of stabilizing the d-c plate current of a temperature-limited thermionic diode noise source to better than 0.02 percent, which corresponds to a noise power stability of better than 0.001 db throughout the current range of 1 ma to 100 ma.

Evaluation of unexpectedly large radiation exposures by means of photographic film,
W. L. McLaughlin, *NBS Tech. Note 161 (Aug. 1962), 15 cents.*
Conventional film types used in personnel monitoring film badges are suitable for measuring X- and γ-radiation exposures only up to 1,000 R. By using special processing procedures, it is possible to extend the range of the less sensitive component of most commercial film packets up to at least 10,000 R. Limitations in precision of readings due to changes in rate dependence, energy dependence, and changes in the shape of the characteristic curve in this range are discussed.

Exchange behavior of kaolins of varying degrees of crystallinity,
W. C. Ormsby, J. M. Shartsis, and K. H. Woodside, *J. Am. Ceram. Soc.* **45**, *No. 8, 361–366 (Aug. 1962).*
Particle-size fractions of several Georgia kaolins, which were prepared by sedimentation procedures, were examined from the standpoint of crystallinity, cation-exchange capacity, and surface area. Crystallinity was studied using X-ray techniques, exchange capacities were measured using the manganese saturation method, and surface areas were determined using glycerol adsorption techniques. A linear relation was obtained between surface areas and exchange capacities when areas were increased by decreasing the particle size or by changing from well crystallized to poorly crystallized kaolins. In most cases, the empirically determined crystallinity ratios indicated a change in crystallinity with change in particle size, the crystallinity generally increasing with decreasing particle size in individual samples. However, unlike the relation noted for exchange, the crystallinity did not consistently correlate with area changes both among the various samples and within the various particle size fractions of a single sample. These results suggest that the relatively high cation exchange capacities of poorly ordered kaolins is more directly a result of high surface areas with crystallinity playing, at most, a very minor role.

Microwave measurements in the NBS Electronic Calibration Center,
R. E. Larson, *Inst. Elec. Engrs.* **109**, *Pt. B, Suppl. No. 23, 644–650 (1962).*
In the Electronic Calibration Center of the National Bureau of Standards Radio Standards Laboratory, Boulder, Colo., work is proceeding towards the establishment or extension of calibration services over a broad range of frequencies in the microwave region. At the present time, measurements can be made over limited portions of the frequency spectrum for the quantities of low-level c.w. power, reflection coefficient, frequency and attenuation. Calibration services are currently provided for all of these quantities. Instrumentation for the measurement of microwave noise is near completion. A survey is given of the microwave measurement techniques employed in this work.

A modulated sub-carrier technique of measuring microwave attenuation,
G. E. Schafer and R. R. Bowman, *Inst. Elec. Engrs.* **109**, *Pt. B, Suppl. No. 23, 783–786 (1962).*
A method of measuring microwave attenuation is proposed which has the advantages of an *i-f* substitution method with single-sideband operation. However, ordinary amplitude modulation is used, and neither the carrier nor one of the sidebands needs to be suppressed.
Two versions of instrumenting this method are described and some operational hints are given. One of these versions is capable of high accuracy with commercially available equipment. The proposed method was tested by measuring the relative attenuation of a microwave variable attenuator at 9.3897 Gcps, attaining a precision of 0.0001 db at 0.01 db and 0.2 db at 50 db. The measurements are compared with calibrations performed by other methods. A special comparison with values obtained from d–c substitution techniques was made in which environmental effects were largely eliminated.
Factors affecting the accuracy of measurements made by this technique are discussed. Some of the precautions necessary to attain high accuracy are given.

A survey of microwave power-measurement techniques employed at the National Bureau of Standards,
G. F. Engen, *Inst. Elec. Engrs.* **109**, *Pt. B, Suppl. No. 23, 734–739 (1962).*
The bolometric technique of power measurement is an important part of the microwave art. The paper describes certain refinements and extensions of this basic method which have been developed at the Boulder Laboratories of the National Bureau of Standards and which provide the basis for a microwave power-calibration service.
The attendant problems may be divided into three categories: (i) measurement of the substituted or bolometric bias power, (ii) evaluation of the d.c. r.f. substitution error, and (iii) determination of the bolometer mount efficiency.
In the first area, the Laboratory has developed precise and automatic d.c. instrumentation which permits an accuracy of $0 \cdot 1$ percent to be realized on a routine basis. In the second and third areas, microcalorimetric techniques enable a determination of the total microwave power dissipated within the bolometer mount; this, when compared with a simultaneous bolometric measurement, determines the combined effect of the substitution error and mount efficiency. Another interesting tool for evaluating bolometer-mount efficiency is provided by the Kerns impedance method. The implementation of this technique has always proved a real

190

challenge, but recent applications of modified reflectometer techniques to the problem have resulted in improved accuracy. Consistent agreement, within a half percent, with microcalorimetric determinations has been realized at X-band frequencies.

Because a bolometer completely absorbs the power being measured, the problem of comparing or calibrating a bolometer mount in terms of a second mount is inherently more difficult than that of comparing two voltmeters or ammeters where simultaneous observations of the same quantity are possible. This problem has, in fact, been a major limitation to the accuracy so far achieved in the art.

Two methods of dealing with this problem have been developed which employ directional couplers and related techniques.

A variable-parameter direct-current switching filter, G. F. Montgomery, *Proc. IRE* **50**, *No. 9, 1986 (Sept. 1962)*.

In a direct-current circuit controlled by a switch, the contact arc is suppressed and the current transient is modified by using a variable-parameter switching filter. A rectifier varies the network structure during contact make and break.

Synthesis of an immittance function with two negative impedance converters, S. B. Geller, *IRE Trans. Circuit Theory* **CT-9**, *No. 3, 291 (Sept. 1962)*.

A technique is shown for synthesizing a function such as $Y(s) = 1$ through the use of two negative impedance con-

 s

verters. Kinariwala's method is used with a limit process when the function fails in one of the required constraints.

The spontaneous martensitic transformations in 18% Cr, 8% Ni steels, R. P. Reed, *Acta Met.* **10**, *865–877 (Sept. 1962)*.

On cooling 18 percent Cr, 8 percent Ni steels transform martensically to two products (ϵ and α'). Sheets representing either ϵ or stacking fault clusters have been observed to form prior to the formation of α'. Photographic sequences demonstrating the formation of α' from these sheets are included. Some transformation characteristics of both ϵ and α' are discussed.

The morphology of the α' has been determined. It was found that the α' formed as long, narrow plates and that these plates were bounded by 111 sheets. The long direction of the plates was parallel or nearly parallel to $<110>$. If they were parallel to $<110>$ then the plates had 225 habit planes. It they deviated from $<110>$ then the habit plane was not (225), possible alternate habit planes are plotted. In addition, the [111] habit plane was observed.

A high speed pyrometer, G. A. Hornbeck, *Book, Temperature, Its Measurement and Control in Science and Industry* **III**, *Pt. 2, 425–428 (Reinhold Publ. Corp., New York, N.Y., 1962)*.

A new high-speed selective spectrometer employed as a multiwavelength pyrometer is described. This instrument is essentially a device which permits the sequential measurement of a number of narrow bands of radiation at any chosen wavelengths at a high rate of speed. The prototype instrument which is described was designed to demonstrate a three wavelength pyrometer with a presentation rate permitting 1,000 temperature determinations per second.

The viscous heating correction for viscometer flows, E. A. Kearsley, *Trans. Soc. Rheology* **VI**, *253–261 (1962)*.

A method is demonstrated for solving simple steady flows of incompressible Newtonian fluid with viscous heating. As an example, a generalization of Poiseuille flow is solved in simple terms.

Wavelengths, energy levels, and pressure shifts in mercury 198, V. Kaufman, *J. Opt. Soc. Am.* **52**, *No. 8, 866–870 (Aug. 1962)*.

The vacuum wavelengths of 27 lines of Hg^{198} and 6 lines of Kr^{86} have been measured relative to the international standard of length, the Kr^{86} line at 6057.80211 Å, by photographic Fabry–Perot interferometry. These measurements were made with Hg^{198} electrodeless lamps containing argon at pressures of $\frac{1}{4}$, 3, and 10 mm Hg and a Kr^{86} hot-cathode lamp operated according to the conditions prescribed by the International Conference on Weights and Measures in 1960. Energy-level values have been derived from the wavelengths of each of the Hg^{198} sources, and on the basis of these values, the energy level and wavelength shifts per unit pressure of argon have been calculated. The suitability of the Hg^{198} electrodeless lamp as a source of wavelength standards for interferometric measurement of length and wavelength is discussed.

A network transfer theorem, G. F. Montgomery, *IRE Trans. Audio* **AU-10**, *No. 3, 88 (May–June 1962)*.

For a linear, passive, reciprocal two-port network, the forward, open-circuit voltage transfer ratio is equal to the reverse, short-circuit current transfer ratio.

Strengthening of hot work die steels, C. R. Irish and S. J. Rosenberg, *Trans. Quart.* **55**, *No. 3, 613–623 (Sept. 1962)*.

A study of four hot-work die steels of the 5 percent chromium type showed that all retained a high percentage of their room temperature strengths at temperatures up to 800 °F.

At 600 °F, the 1,000-hour stress-rupture life was in excess of 98 percent of the short-time tensile strength at that temperature. At 800 °F, failures were obtained at stresses between 85 and 98 percent of the short-time strength.

Specimens that survived 1,000 hours in the stress-rupture machines were subsequently tensile tested at room temperature. The results obtained indicated that the strength of these specimens had been significantly increased.

Acoustical interferometer employed as an instrument for measuring low absolute temperatures, G. Cataland and H. H. Plumb, *J. Acoust. Soc. Am.* **34**, *No. 8, 1145–1146 (Aug. 1962)*.

Values of absolute temperatures at 2° and 20 °K have been determined from experimental measurements of the speed of sound as a function of pressure in helium gas. The acoustical interferometer was the instrument employed in the measurements, and the accuracy achieved in the experiment indicates that sonic thermometry at low temperatures may be competitive with other conventional thermometry techniques.

Correlation of factors influencing the pressures generated in multi-anvil devices, J. C. Houck and U. O. Hutton, *Am. Soc. Mech. Engrs. Paper 62–WA–254 (1962)*.

Tests were performed with three different multianvil wedge-type high pressure devices, using pyrophyllite as the sample holder. Two devices made use of tetrahedrons, nominally $\frac{1}{2}''$ and $1''$ on an edge; the third used a nominal $\frac{3}{8}''$ cube. The change in electrical resistance was used to detect the transitions of bismuth I–II (25.2 kb) bismuth II–III (26.6 kb) and barium (59 kb). Major effects observed were: (1) Oven drying the pyrophyllite sample holders to remove moisture gave significantly lower anvil forces to reach the transition. (2) A silver chloride sleeve not only caused transition to go to completion for smaller increases in applied anvil forces, but also reduced the anvil forces required to reach the bismuth II–III and the barium (59 kb) transition pressures. (3) A wide range of sizes of sample holders in the same size die had little effect on anvil force required to reach the transition. (4) Comparison of results with the $\frac{1}{2}''$ and $1''$ tetrahedrons showed that the ram loads required to attain the transitions were proportional to the face areas of the anvils.

A "two-stage" device was constructed by insertion of hardened steel truncated cones in the faces of the pyrophyllite tetrahedrons. This arrangement permitted the attainment of the bismuth 88 kb transition with the ram load reduced to about one-half of that required for the single-stage arrangement.

The temperature dependence of flow and fracture characteristics of an age-hardenable alloy, W. D. Jenkins and W. A. Willard, *Trans. ASM,* **55**, *No. 3, 580–598 (Sept. 1962)*.

Mechanisms contributing to flow, fracture and ductility of polycrystalline Duranickel tensile specimens tested in the temperature range 75° to 1,200 °F are discussed. The temperature dependence of the yield point phenomenon and reversals in strength-temperature curves of the annealed metal is attributed to precipitation of Ni_3Al during deformation. Increase in strength and decrease in ductility due to aging are rationalized on the basis of the presence of precipitates

which interfere with the motion of dislocations. The density, distribution, shape and size of slip bands, precipitated particles, cracks and cavities in the specimens before and after fracture were observed by means of optical and electron microscopy and are discussed by use of dislocation theory. The influence of aging on tensile deformation and of tensile deformation on aging is partially analyzed by means of hardness values.

The effect of experimental variable including the martensitic transformation on the low-temperature mechanical properties of austenitic stainless steels, C. J. Guntner and R. P. Reed, *Trans. ASM* **55,** *No. 3,399–491 (Sept. 1962).*
The austenitic stainless steels in general are excellent low temperature structural materials. Most of these steels undergo a martensitic transformation which produces both a hexagonal close-packed phase and a body centered phase and affects the mechanical properties.
Tensile, notched-tensile and impact tests have been conducted on AISI 202, AM 350, USS Tenelon, AISI 310 and 5 commercial grades of AISI 304 and AISI 304L at temperatures between 300 and 4° K. Values are obtained for tensile strength, notched-tensile strength, impact strength, tensile elongation, tensile reduction of area and notched-tensile-reduction of area. A plot is also included on the percent phase transformation as a function of tensile strain and temperature. Experimental variables considered are temperature, strain rate, specimen geometry and initial microstructure. The influences of the experimental variables and the martensitic transformation characteristics on the mechanical properties are discussed.

Characteristics of resistance strain gages, R. L. Bloss, *Book, Semiconductor and Conventional Strain Gages, Ed. Mills Dean, III and R. D. Douglas, chapt. VII, 123–142 (Academic Press, Inc., New York, N.Y., Oct. 1962).*
Although resistance strain gages are very useful devices for many applications, their characteristics and limitations must be examined closely when use in a new situation is contemplated. The characteristics and factors which may limit the usefulness of these gages include (1) strain sensitivity, (2) temperature sensitivity, (3) resistance instability, (4) shelf life of components, (5) effects of moisture, (6) incompatibility of components, (7) fatigue life, (8) frequency range, (9) magnetostrictive effects, and (10) incompatibility with environment. These factors are discussed and illustrated.

Standard tests for electrical properties, A. H. Scott, *SPE J. 1375–1378 (Nov. 1962).*
A discussion is given of the use of standard tests to determine (1) volume and surface resistivity, (2) permittivity (dielectric constant) and dissipation factor, (3) dielectric strength and (4) arc resistance or tracking. Both American and International Tests are cited.

Chromium plating by thermal decomposition of dicumene chromium, W. H. Metzger, Jr., *Plating* **49,** *No. 11, 1176 (Nov. 1962).*
A technical note describing experiments on chromium plating by thermal decomposition of dicumene chromium.

New wavemeter for millimeter wavelengths, R. W. Zimmerer, *Rev. Sci. Instr.* **33,** *No. 8, 858–859 (Aug. 1962).*
A new wavemeter of simple design is described. The principle of operation makes use of a new development in physical optics. The actual performance of the device was measured and compared with the theory.

The use of a Venturi tube as a quality meter, R. V. Smith, P. C. Wergin, J. F. Ferguson, and R. B. Jacobs, *J. Basic Eng.* **84,** *411–412 (Sept. 1962).*
It is shown that the relationship between the pressure drop Δp, the mass rate of flow m, and the quality x, of the two-phase fluid flowing through a Venturi can be correlated by the following expression

$$\frac{\sqrt{\Delta p}}{m} = a + bx$$

It follows that if the pressure drop and mass flow rate are measured, the quality is easily computed.

Dislocation loops in deformed copper, A. W. Ruff, Jr., *Fifth Intern. Congress for Electron Microscopy, p. J–10 (Academic Press, Inc., New York, N.Y., 1962).*
An examination by transmission electron microscopy of single crystal copper foils which were deformed 12 percent and 20 percent by rolling, has revealed the presence of considerable numbers of small dislocation loops. Average values are given for the dislocation density and the loop density. It is believed that these loops were formed from point defects generated during the deformation, and that the point defect concentration immediately after deformation was at least 10^{-5}. The diffraction contrast effects associated with the loops indicate that the dislocations are complete and not partial.

Applications of resistance thermometers to calorimetry, G. T. Furukawa, *Book, Temperature, Its Measurement and Control in Science and Industry* **III,** *Pt. 2, 317–328 (Reinhold Publ. Co., New York, N.Y., 1962).*
The importance of the resistance thermometer in the accurate measurement of both temperature and the heat leak of the calorimeter is discussed. The final accuracy of the determination of heat capacity is shown to be dependent upon the accurate and consistent measurement of heat input to the sample and the corresponding rise in temperature. The various heat-capacity calorimeters used in the range from 10 to 400 °K are briefly described with emphasis upon the applications of resistance thermometers, the methods for calibrating them, and the problems associated with the design of calorimeter vessels. Comparison is made of the thermometric properties of platinum, copper, indium, lead and gold-silver alloy. The need for high relative accuracy in the measurement of ΔR is emphasized. The various temperature scales used in calorimetry are compared and their applications are described. Absorption spectrum of carbon vapor in solid argon at 4° and 20 °K, R. L. Barger and H. P. Broida, J. Chem. Phys. **37,** No. 5, 1152–1153 (Sept. 1, 1962).

Obtaining the internal junction characteristics of a transistor for use in analog simulation, S. B. Geller, *IRE Trans. Electron. Computers* **EC–11,** *No. 5, 709–710 (Oct. 1962).*
A technique is described for making the internal base-to-emitter junction characteristics of an alloy junction transistor available to an analog computer simulation process. This is accomplished with an active feedback network that continuously compensates for the internal voltage drop across the extrinsic base-spreading resistance at all base current levels.

The thermodynamic scale of temperature below 1 °K, R. P. Hudson, *Book, Temperature, Its Measurement and Control in Science and Industry* III, *Pt. I, 51–57 (Reinhold Publ. Co., New York, N.Y., 1962).*
Following a brief discussion of the principles of magnetic thermometry, a description is given of the main methods use to derive the relation between the "magnetic scale" and the absolute scale of temperature. Experimental results published since 1953 are summarized. An account is given of the measurement of absolute temperature using the anisotropy of radiation emitted from oriented radioactive nuclei. Recent work on the intercomparison of "nuclear orientation scales", and on the comparison of one such scale with a magnetic scale, is reviewed. There follows a short account of the adaptation of the magnetic-cooling method to nuclear paramagnetics and the production of temperatures of the order of one microdegree Kelvin.

Precision phase meter, D. M. Waters, D. Smith, and M. C. Thompson, Jr., *IRE Trans. Instr.* **I–II,** *64–66 (Sept. 1962).*
A precision electromechanical phase meter has been developed to record slow, continuous phase variations often encountered in radio propagation research. The phase meter will follow phase variations up to several complete cycles unambiguously and small phase variations as fast as 1 c/s.

The effect of temperature and humidity on the oxidation of air-blown asphalts, P. G. Campbell, J. R. Wright, and P. B. Bowman, *Mater. Res. Std.* **2,** *No. 12, 988–995 (Dec. 1962).*
The effects of temperature and humidity on the oxidation of air-blown roofing asphalts were determined by measuring the changes in infrared absorption in the carbonyl band caused by carbon-arc exposure of the asphalts under varying conditions

of temperature and relative humidity. Asphalt oxidation was measured for both fixed periods of exposure, and as a function of exposure time. Both temperature and humidity affected the rate of asphalt oxidation, with temperature being the more critical parameter. The sensitivity of the asphalts to changes in temperature and relative humidity varied with asphalt source and asphalt durability. The relative order of oxidation stability of a series of asphalts exposed outdoors was, in general, the same as that obtained with exposure to carbon-arc radiation. The formation and subsequent decomposition of an ashpalt-oxygen-water complex is proposed as a possible mechanism for the effects of temperature and relative humidity on asphalt oxidation rates.

The thermodynamic temperature scale, its definition and realization, C. M. Herzfeld, *Book, Temperature, Its Measurement and Control in Science and Industry III, Pt. I, 41–50 (Reinhold Publ. Co., New York, N.Y., 1962).*
A temperature scale based on thermodynamics is conceptually straightforward. The usual definition is given and the realization of the scale by means of gas thermometry is discussed.
The scale can be extended by appeal to the statistical mechanical interpretation of thermodynamics. Ways of doing this for high and low temperatures, using radiation and magnetic methods respectively, are presented.
Statistical mechanical arguments make possible the use of the concept of temperature for systems differing greatly from those contemplated in classical thermodynamics.

Use of a "peek-a-boo" information retrieval technique for a personal reference file, J. A. Bennett, *J. Wash. Acad. Sci.* **52**, *No. 9, 216–219 (Dec. 1962).*
Optical coincidence subject cards have many advantages for indexing a personal reference file. A system having a capacity of 1,500 items has proven very useful and does not require complicated punching and searching equipment.

Photographic strain measuring technique for use above 3,000 F. L. Mordfin and T. Rubusto, Jr., *Proc. Instr. Soc. Am.* **17**, *Pt. 1, 3.4.62–1 (1962).*
A technique for measuring local surface strains in a structural test specimen is proposed, in which gage point markings applied to the surface of the specimen are photographed. This paper describes an exploratory application of this method to the measurement of axial and lateral strains in the tensile test of a molybdenum rod at 3,500 F in a vacuum.

The speed of processes involved in electroplating: movement of solute, attainment of the steady state and formation of metal, A. Brenner, *49th Annual Tech. Proc. Am. Electroplaters Soc. p. 9–13 (1962).*
The time involved in various processes occurring during ordinary electroplating was discussed. Ions moved to the cathode at the rate of 10^{-4} cm/sec. The upward movement of convection currents along an electrode was about 6 cm per minute. The time required to reach a steady state of electrolysis was about 2 minutes. Deposition of metal can be made to occur with a microsecond pulse of current. By means of galvanostagometry is was shown that an electrode reaction occurs in less than 5 microseconds after a circuit is closed.

Journal of Research 67A (Phys. and Chem.), No. 1 (Jan.–Feb. 1963), 70 cents.

Heat of formation of calcium aluminate monosulfate at 25 °C. H. A. Berman and E. S. Newman.
2,3-Dimethylpentane and 2-methylhexane as a test mixture for evaluating highly efficient fractionating columns. E. C. Kuehner.
Phase equilibrium relations in the Sc_2O_3-Ga_2O_3 system. S. J. Schneider and J. L. Waring.
Analysis of two infrared bands of CH_2D_2. Wm. B. Olson, H. C. Allen, Jr., and E. K. Plyler.
Precise coulometric titrations of halides. G. Marinenko and J. K. Taylor.
Radial distribution study of vitreous barium borosilicate. G. J. Piermarini and S. Block.

Dynamic compressibility of poly (vinyl acetate) and its relation to free volume. J. E. McKinney and H. V. Belcher.
An investigation of the constitution of the mercury-tin system. D. F. Taylor and C. L. Burns.
Effect of methyl bromide additions on the flame speed of methane. C. Halpern.

Journal of Research 67A (Phys. and Chem.), No. 2 (Mar.–Apr. 1963), 70 cents.

Third spectrum of palladium (Pd III). A. G. Shenstone.
Broadening of the rotational lines of carbon monoxide by HCl and argon. R. J. Thibault, J. H. Jaffe, and E. K. Plyler.
Theory of frustrated total reflection involving metallic surfaces. T. R. Young and B. D. Rothrock.
Quantitative metallography with a digital computer: application to a Nb-Sn superconducting wire. G. A. Moore and L. L. Wyman. (See above abstracts.)
Moiré fringes produced by a point projection X-ray microscope. S. B. Newman. (See above abstracts.)
Cyclic polyhydroxy ketones. I. Oxidation products of hexahydroxybenzene (benzenehexol). A. J. Fatiadi and H. S. Isbell.
Effect of pressure and temperature on the refractive indices of benzene, carbon tetrachloride, and water. R. M. Waxler and C. E. Weir.
Pressure-density-temperature relations of fluid para hydrogen from 15 to 100 °K at pressures to 350 atmospheres. R. D. Goodwin, D. E. Diller, H. M. Roder, and L. A. Weber.
A method for determining the elastic constants of a cubic crystal from velocity measurements in a single arbitrary direction; application to $SrTiO_3$. J. B. Wachtman, Jr., M. L. Wheat, and S. Marzullo. (See above abstracts.)

Journal of Research 67B (Math. and Math. Phys.), No. 1 (Jan.–Mar. 1963), 75 cents.

Evaluation of a generalized elliptic-type integral. L. F. Epstein and J. H. Hubbell.
An algorithm for obtaining an orthogonal set of individual degrees of freedom for error. J. M. Cameron.
Recognition of completely mixed games. A. J. Goldman.
A new type of computable inductor. C. H. Page. (See above abstracts.)
Numerical computation of the temporal development of currents in a gas discharge tube. W. Börch-Supan and H. Oser.
Tables of genera of groups of linear fractional transformations. H. Fell, M. Newman, and E. Ordman.

Journal of Research 67D (Radio Prop.), No. 1 (Jan.–Feb. 1963), 70 cents.

A lunar theory reasserted—a rebuttal. J. V. Evans. Point-to-point communication on the moon. L. E. Vogler.
HF communication during ionospheric storms. G. E. Hill.
Use of surface refractivity in the empirical prediction of total atmospheric refraction. W. R. Iliff and J. M. Holt.
Effective sunspot numbers. W. B. Chadwick.
On the theory of radio wave propagation over inhomogeneous earth. K. Furutsu.
Fields of electric dipoles in sea water (a correction). W. Anderson.
Composition of reflection and transmission formulae. J. Heading.
Titheridge coefficients for the polynomial method of deducing electron density profiles from ionograms. A. R. Long and J. O. Thomas.
Input admittance of linear antennas driven from a coaxial line. T. T. Wu. (See above abstracts.)

Journal of Research 67D (Radio Prop.), No. 2 (Mar.–Apr. 1963), 70 cents.

The protection of frequencies for radio astronomy. R. L. Smith-Rose.
Radar reflections from the moon at 425 Mc/s. G. H. Millman and F. L. Rose.

Sunset and sunrise in the ionosphere: effects on the propagation of longwaves. J. Rieker.

Correction of atmospheric refraction errors in radio height finding. W. B. Sweezy and B. R. Bean.

Empirical determination of total atmospheric refraction at centimeter wavelengths by radiometric means. A. C. Anway.

Propagation of radiofrequency electromagnetic fields in geological conductors. V. Fritsch, translated from German by A. P. Barsis.

WWV reception in the arctic during ionospheric disturbances. G. E. Hill and J. R. Herman.

Height-gain for VLF radio waves. J. R. Wait and K. P. Spies.

Perturbation method in a problem of waveguide theory. D. Fox and W. Magnus.

Some wave functions and potential functions pertaining to spherically stratified media. C. T. Tai.

Radiation from a plasma-clad axially-slotted cylinder. W. V. T. Rusch.

Two- and three-loop superdirective receiving antennas. E. W. Seeley.

Hallen's method in the problem of a cavity-backed rectangular slot antenna. J. Galejs.

Relative convergence of the solution of a doubly infinite set of equations. R. Mittra.

———

Periodicals received in the Library of the National Bureau of Standards, July 1962, N. J. Hopper, NBS Mono. 57 (Nov. 23, 1962), 25 cents (Supersedes NBS Circular 563 and the 1st supplement to NBS Circular 563).

Handbook for CRPL Ionospheric Predictions Based on Numerical Methods of Mapping, S. M. Ostrow, NBS Handb. 90 (Dec. 21, 1962), 40 cents (Supersedes Circ. 465).

Report of the 47th National Conference on Weights and Measures 1962, NBS Misc. Publ. 244 (Nov. 23, 1962), 75 cents.

Hydraulic research in the United States 1962, H. K. Middleton, NBS Misc. Publ. 245 (Oct. 26, 1962), $1.00.

1962 Research Highlights of the National Bureau of Standards, Annual Report, NBS Misc. Publ. 246 (Dec. 1962), 70 cents.

Quarterly radio noise data, March, April, May 1962 and corrigendum for Technical Notes 18–1 through 18–11, W. Q. Crichlow, R. T. Disney and M. A. Jenkins, NBS Tech. Note 18–14, (Aug. 9, 1962) 50 cents.

Mean electron density variations of the quiet ionosphere, No. 8—October 1959, J. W. Wright, L. R. Wescott, and D. J. Brown, NBS Tech. Note 40–8, (Sept. 1962), 35 cents.

Synoptic radio metrology, B. R. Bean, J. D. Horn, and L. P. Riggs, NBS Tech. Note 98, (Oct. 1962), 50 cents.

Bibliography on direction finding and related ionospheric propagation topics, 1955–1961, O. D. Remmler, NBS Tech. Note 127, (Oct. 1962), 60 cents.

Equatorial spread, F. W. Calvert, NBS Tech. Note 145 (Aug. 1, 1962), 60 cents.

The energy parameter B for strong blast waves, D. L. Jones, NBS Tech. Note 155, (July 1962), 25 cents.

Thermal balance in the F region of the atmosphere, D. C. Hunt, NBS Tech. Note 162 (Sept. 1962), 50 cents.

Spectrophotometric determination of hydroperoxide in diethyle ether, W. C. Wolfe, Anal. Chem. **34**, No. 10, 1328–1330 (Sept. 1962).

Analysis of the hydroxyl radical vibration rotation spectrum between 3900 A and 15000 A, A. M. Bass and D. Garvin, J. Mol. Spectry. **9**, No. 2, 114–123 (Aug. 1962).

Structure and structure imperfections of solid β-oxygen, E. M. Horl, Acta Cryst. **15**, No. 9, 845–850 (Sept. 1962).

Kinetics of Cs^+ desorption from tungsten, M. D. Scheer and J. Fine, J. Chem. Phys. **37**, No. 1, 107–113 (July 1962).

Effect of additives on silver iodide particles exposed to light, G. Burley and D. W. Herrin, J. Appl. Meteorol. **1**, No. 3, 355–356 (Oct. 1962).

Symposium on teaching of materials: Aspects of material behavior significant to engineers, J. M. Frankland, J. Eng. Mech. Div.–Proc. Am. Soc. Civil Engr. Paper No. 3246, 75–81 (Aug. 1962).

Neutral meson photoproduction from complex nuclei, R. A. Schrack, J. E. Leiss and S. Penner, Phys. Rev. **127**, No. 5, 1772–1783 (Sept. 1, 1962).

Excitation of modes at very low frequency in the earth-ionosphere wave guide, J. R. Wait, J. Geophys. Res. **67**, No. 10, 3823–3828 (Sept. 1962).

The growth and roots of electronic computers, J. L. Little, Student Quart. and EE Digest, Joint Publ. of Am. Inst. Elec. Engr. and IRE **1**, No. 1, 64–70 (Sept. 1962).

Acoustical interferometer employed as an instrument for measuring low absolute temperatures, G. Cataland and H. H. Plumb, J. Acoust., Soc. Am. **34**, No. 8, 1145–1146 (Aug. 1962).

The Hall effect, D. O. Webster, Student Quart. and EE Dig. pp. 82–86 (Sept. 1962).

Dimensional changes in complete dentures on drying, wetting, and heating in water, G. C. Paffenbarger and W. T. Sweeney, J. Am. Dental Assoc. **65**, No. 4, 495–505 (Oct. 1962).

The vapor phase radiolysis of propane-d_8 in the presence of other hydrocarbons, S. G. Lias and P. Ausloos, J. Chem. Phys. **37**, No. 4, 877–883 (Aug. 15, 1962).

A note on the propagation of electromagnetic pulses over the earth's surface, J. R. Wait, Can. J. Phys. **40**, 1264–1269 (1962).

Surface behavior of silver single crystals in fused sodium chloride, J. Kruger and K. H. Stern, J. Electrochem. Soc. **109**, No. 10, 889–894 (Oct. 1962).

Rotational perturbations in CN I. Zero-field theory, optical Zeeman effect, and microwave transition probabilities, H. E. Radford and H. P. Broida, Phys. Rev. **128**, No. 1, 231–242 (Oct. 1962).

A new ionospheric multipath reduction factor (MRF), R. K. Salaman, IRE Trans. Commun. Systems **CS–10**, 220–222 (June 1962).

The effect of electrolytes on the sodium chromotropism of bix-(meso-2,3-diaminobutane)-nickel (II) ions, D. L. Leussing, J. Harris, and P. Wood, J. Phys. Chem. **66**, 1544–1546 (Aug. 1962).

The photodisintegration of bismuth and the lead isotopes, E. G. Fuller and E. Hayward, Nuclear Phys. **33**, 431–448 (1962).

Comparison of the relative acidic strengths of the isomeric dinitrophenols in benzene and water, M. M. Davis, J. Am. Ceram. Soc. **84**, 3623–3627 (1962).

The basis of the measurement system, A. V. Astin, Proc. IRE **50**, No. 5, 614–616 (May 1962).

Predicting the performance of long distance tropospheric communication circuits, A. P. Barsis, K. A. Norton, and P. L. Rice, IRE Trans. Commun. Systems **CS–10**, 2–22 (Mar. 1962).

A neuron model which performs analog functions, D. R. Boyle, Proc. San Diego Symp. Biomedical Engr., San Diego, Calif., pp. 256–263 (June 19–21, 1962).

The system of Bi_2O_3–B_2O_3, E. M. Levin and C. L. McDaniel, J. Am. Ceram. Soc. **45**, No. 8, 355–360 (Aug. 1962).

Electron energy losses in solids and their influence on the electron diffraction diagram, L. Marton, J. Phys. Soc. Japan **17**, Suppl. B–II, 68–73 (1961).

Scattered radiation from large Cs^{137} sources, L. Costrell, Health Phys. **8**, No. 5, 491–498 (Oct. 1962).

Optical detection of microwave transitions in electronically excited CN produced by a chemical reaction, R. L. Barger, H. P. Broida, A. J. Estin, and H. E. Radford, Phys. Rev. Letters **9**, No. 8, 345–346 (Oct. 15, 1962).

Sum rules for vibrational-rotational energy levels including centrifugal distortion, H. C. Allen, Jr., and W. B. Olson, J. Chem. Phys. **37**, No. 2, 212–214 (July 15, 1962).

Optical constants of aluminum in vacuum ultraviolet, R. LaVilla and H. Mendlowitz, Phys. Rev. Letters **9**, No. 4, 149–150 (Aug. 15, 1962).

Microwave spectrum of methyldifluoroarsine, L. J. Nugent and C. D. Cornwell, J. Chem. Phys. **37**, No. 3, 523–534 (Aug. 1962).

The industrial significance of National Bureau of Standards research, G. M. Kline, Mod. Plastics **40**, No. 2, 149 (Oct. 1962).

Comments on "A new precision low level bolometer bridge," G. F. Engen, Proc. IRE **50**, No. 9, 1997 (Sept. 1962).

Solar particles in interplanetary space, C. S. Warwick, Sky and Telescope **XXIX**, No. 3, 133–136 (Sept. 1962).

The kinetics of the heat precipitation of collagen, J. M. Cassel, L. Mandelkern, and D. E. Roberts, J. Am. Leather Chem. Assoc. **LVII**, No. 11, 556–575 (Nov. 1962).

Oxidation of high polymers, M. Tryon and L. A. Wall, Book, Autoxidation and Antioxidants II, Chapt. 19, p. 919 (Interscience Publ. Inc., New York, N.Y., 1962).

Vibrational-rotational spectroscopy, H. C. Allen, Jr., and W. B. Olson, Ann. Rev. Phys. Chem. **13**, 221–240 (1962).

Realistic diatomic potential function, H. W. Wooley, J. Chem. Phys. **37**, No. 6, 1307–1316 (Sept. 1962).

Absorption spectrum of carbon vapor in solid argon at 4° and 20° K, R. L. Barger and H. P. Broida, J. Chem. Phys. **37**, No. 5, 1152–1153 (Sept. 1, 1962).

Rotation-vibration spectrum of matrix-isolated hydrogen chloride, L. J. Schoen and D. E. Mann, J. Chem. Phys. **37**, No. 5, 1146–1147 (Sept. 1, 1962).

Very-low-frequency phase observations on the ionospheric effects of the solar flare of September 28, 1961, A. G. Jean and J. H. Crary, J. Geophys. Res. **67**, No. 12, 4903–4905 (Nov. 1962).

The radiant energy from sources in the far infrared, E. K. Plyler, D. Yates, and H. A. Gebbie, J. Opt. Soc. Am. **52**, No. 8, 859–861 (Aug. 1962).

Graphical evaluation of analytical results, W. J. Youden, Proc. Conf. Chemical Control Problems, pp. 1–15 (Spon. Natl. Plant Food Inst., Washington, D.C., 1959).

The measurement of ionospheric drifts by means of a Doppler technique, K. Davies, J. Geophys. Res. **67**, No. 12, 4909–4913 (Nov. 1962).

Laboratory astrophysics, L. M. Branscomb and R. N. Thomas, Phys. Today **15**, 42–44, 46 (Nov. 1962).

Standards—The key to success in engineering, A. T. McPherson, Rev. Centro Estud. Ing. **130**, No. 656, 352–355 (1961).

Nuclear photoeffect in deformend nuclei, E. Hayward, Rendiconti Della Scuola Intern. Fisica "Enrico Fermi" **XV**, 214–222 (Academic Press Inc., New York, N.Y., 1962).

Standards for the future, R. D. Huntoon, Proc. Annual Spring Conf. Standards Engineers Society, Boston Section and Company Member Conf. of American Standards Association, May 24–25, 1962 (Boston, Mass., 1962).

Intramolecular rearrangements in the solid phase photolysis of 4-methyl-2-hexanone and sec-butyl acetate, R. E. Rebbert and P. Ausloos, J. Chem. Phys. **37**, No. 5, 1158–1159 (Sept. 1, 1962).

The vapor phase fluorescence and its relationship to the photolysis of propionaldehyde and the butyraldehydes, R. P. Borkowski and P. Ausloos, J. Am. Chem. Soc. **84**, No. 21, 4044–4048 (Nov. 5, 1962).

Relaxation of nonequilibrium distributions, K. E. Shuler, Book, Temperature, Its Measurements and Control in Science and Industry III, Pt. I, 27–34 (Reinhold Publ. Co., New York, N.Y., 1962).

Crosslinking of polymers induced by excited species, L. A. Wall and R. B. Ingalls, J. Polymer Sci. Letter to Editor **62**, No. 173, S5–S6 (1962).

Standard electromotive force of the cell H_2; $HBr(m)$; $AgBr$; Ag from 0 to 50°, H. B. Hetzer, R. A. Robinson, and R. G. Bates, J. Phys. Chem. **66**, No. 8, 1423–1426 (1962).

Vacuum ultraviolet photochemistry. IV. Photolysis of propane, H. Okabe and J. R. McNesby, J. Chem. Phys. **37**, No. 6, 1340–1346 (Sept. 1962).

Interpretation of rate experiments with resolved quantum levels, T. Carrington, Discussions Faraday Soc., No. **33**, 44–51 (1962).

Hyperfine structure and intercombination line intensities in the spectra of magnesium, zinc, cadmium, and mercury, R. H. Garstang, J. Opt. Soc. Am. **52**, No. 8, 845–851 (Aug. 1962).

Conformation and frictional properties of polystyrene in dilute solutions, D. McIntyre, A. Wims, L. C. Williams, and L. Mandelkern, J. Phys. Chem. **66**, 1932–1940 (Oct. 1962).

Complex formation of monomeric amides with lithium perchlorate, A. F. Diorio, E. Lippincott, and L. Mandelkern, Nature **194**, No. 4848, 1296–1297 (Sept. 1962).

Absorption bands of carbon dioxide from 2.8 to 4.2 μ, E. K. Plyler, E. D. Tidwell, and W. S. Benedict, J. Opt. Soc. Am. **52**, No. 9, 1017–1022 (Sept. 1962).

Synthesis of higher ketoses by aldol reactions. I. Three D-heptuloses, R. Schaffer, and H. S. Isbell, Org. Chem. **27**, 3268–3270 (Mar. 1962).

Publications for which a price is indicated may be obtained from the Superintendent of Documents, U.S. Government Printing Office, Washington 25, D.C. (foreign postage one-fourth additional). Reprints from outside journals and the NBS Journal of Research may often be obtained from the authors.