Section 11.3
FILE COPY
Do not remove

NATIONAL BUREAU OF STANDARDS REPORT

1600

THE RELIABILITY OF MEASURED VALUES —
FUNDAMENTAL CONCEPTS

by

Churchill Eisenhart

NBS

U. S. DEPARTMENT OF COMMERCE
NATIONAL BUREAU OF STANDARDS

U. S. DEPARTMENT OF COMMERCE
Charles Sawyer, Secretary

NATIONAL BUREAU OF STANDARDS
A. V. Astin, Acting Director

THE NATIONAL BUREAU OF STANDARDS

The scope of activities of the National Bureau of Standards is suggested in the following listing of the divisions and sections engaged in technical work. In general, each section is engaged in specialized research, development, and engineering in the field indicated by its title. A brief description of the activities, and of the resultant reports and publications, appears on the inside of the back cover of this report.

1. **ELECTRICITY.** Resistance Measurements. Inductance and Capacitance. Electrical Instruments. Magnetic Measurements. Electrochemistry.

2. **OPTICS AND METROLOGY.** Photometry and Colorimetry. Optical Instruments. Photographic Technology. Length. Gage.

3. **HEAT AND POWER.** Temperature Measurements. Thermodynamics. Cryogenics. Engines and Lubrication. Engine Fuels.

4. **ATOMIC AND RADIATION PHYSICS.** Spectroscopy. Radiometry. Mass Spectrometry. Physical Electronics. Electron Physics. Atomic Physics. Neutron Measurements. Nuclear Physics. Radioactivity. X-Rays. Betatron. Nucleonic Instrumentation. Radiological Equipment. Atomic Energy Commission Instruments Branch.

5. **CHEMISTRY.** Organic Coatings. Surface Chemistry. Organic Chemistry. Analytical Chemistry. Inorganic Chemistry. Electrodeposition. Gas Chemistry. Physical Chemistry. Thermochemistry. Spectrochemistry. Pure Substances.

6. **MECHANICS.** Sound. Mechanical Instruments. Aerodynamics. Engineering Mechanics. Hydraulics. Mass. Capacity, Density, and Fluid Meters.

7. **ORGANIC AND FIBROUS MATERIALS.** Rubber. Textiles. Paper. Leather. Testing and Specifications. Organic Plastics. Dental Research.

8. **METALLURGY.** Thermal Metallurgy. Chemical Metallurgy. Mechanical Metallurgy. Corrosion.

9. **MINERAL PRODUCTS.** Porcelain and Pottery. Glass. Refractories. Enameled Metals. Building Stone. Concreting Materials. Constitution and Microstructure. Chemistry of Mineral Products.

10. **BUILDING TECHNOLOGY.** Structural Engineering. Fire Protection. Heating and Air Conditioning. Exterior and Interior Coverings. Codes and Specifications.

11. **APPLIED MATHEMATICS.** Numerical Analysis. Computation. Statistical Engineering. Machine Development.

12. **ELECTRONICS.** Engineering Electronics. Electron Tubes. Electronic Computers. Electronic Instrumentation.

13. **ORDNANCE DEVELOPMENT.** Mechanical Research and Development. Electromechanical Fuzes. Technical Services. Missile Fuzing Research. Missile Fuzing Development. Projectile Fuzes. Ordnance Components. Ordnance Tests. Ordnance Research.

14. **RADIO PROPAGATION.** Upper Atmosphere Research. Ionospheric Research. Regular Propagation Services. Frequency Utilization Research. Tropospheric Propagation Research. High Frequency Standards. Microwave Standards.

15. **MISSILE DEVELOPMENT.** Missile Engineering. Missile Dynamics. Missile Intelligence. Missile Instrumentation. Technical Services. Combustion.

NBS PROJECT

15 April 1952

NBS REPORT

1100-53-1199

1600

THE RELIABILITY OF MEASURED VALUES — FUNDAMENTAL CONCEPTS*

by

Churchill Eisenhart

Statistical Engineering Laboratory

NBS

*Talk presented in the Symposium on Precision, Accuracy and Statistical Method, before the American Society of Photogrammetry, Washington, D. C., 10 January 1952.

This report is issued fc , reprinting, or reproduction in any form, either in wh ission in writing is obtained from the Office of the Di hington 25, D. C.

Approved for public release by the Director of the National Institute of Standards and Technology (NIST) on October 9, 2015

FOREWORD

This is the text of an invited address, "The reliability of measured values — fundamental concepts," presented by Dr. Churchill Eisenhart, Chief of the Statistical Engineering Laboratory (Section 3 of Division 11, Applied Mathematics) of the National Bureau of Standards, in the Symposium on Precision, Accuracy, and Statistical Method, sponsored by the American Society of Photogrammetry as part of the program of its 18th Annual Meeting, held in Washington, D. C., on 9-11 January 1952.

Included here also (pp. 33-43) are comments on Dr. Eisenhart's address by Mr. Amrom H. Katz (Chief Physicist, Photographic Laboratory, Air Materiel Command, Wright-Patterson Air Force Base, Dayton, Ohio), who organized and served as chairman of the Symposium, and by Captain Oliver S. Reading (Chief, Division of Photogrammetry, U. S. Coast and Geodetic Survey, Washington 25, D. C.), together with Dr. Eisenhart's replies to the questions raised.

This material will be published in due course in the American Journal of Photogrammetry as a part of the Proceedings of the Symposium.

J. H. Curtiss
Chief, National Applied
Mathematics Laboratories

A. V. Astin
Acting Director
National Bureau of Standards

THE RELIABILITY OF MEASURED VALUES - FUNDAMENTAL CONCEPTS

Table of Contents

Today I plan to discuss some fundamental concepts that have to do with measurement. You will notice that my remarks are highly statistical because I discovered some years ago that the theory of measurement is intimately tied up with statistical concepts and methods. It was as an undergraduate major in mathematical physics at Princeton University that I first began to give serious attention to the theory of measurement, and especially to that part of the subject known as the "theory of errors." I had done a little reading on the theory of errors in connection with my undergraduate courses in physics, had found the subject rather dull, and very likely would have given it no further consideration had it not been for the influence of Dr. Edward U. Condon, then

(1930-1937) associate professor of physics at Princeton, more recently Director of the National Bureau of Standards (1945-1951), and now Director of Research and Development at the Corning Glass Works.

Dr. Condon, learning from one of my other professors that I had shown some interest in the theory of probability, suggested that I take a second look at the theory of errors after familiarizing myself with some of the more recent developments in statistical theory and methodology. To punctuate his suggestion he loaned me his personal copy of a then little known book by R. A. Fisher, Statistical Methods for Research Workers (Edinburgh: Oliver & Boyd, 1st ed., 1925; 11th ed., 1950). After perusing this volume for quite some time I returned to Dr. Condon with (a) a confession to the effect that I had been unable to determine the mathematical basis of much that I had read in this book, and (b) a conviction that, assuming this book to be sound, it carried a great message to experimental physicists and chemists who conduct and interpret experiments involving only a small number of observations and that a "translation" of its message should be made available to physical scientists without delay. Dr. Condon replied that, with regard to the difficulty experienced in fathoming the mathematical basis of the book, I was not alone; that he believed the book to be sound; and "How about my undertaking the 'translation'?" That was the beginning. As my first effort, I wrote, in my junior year at Princeton, an essay entitled "A Discussion of 'Student's' Method for Testing the Significance of a Small Number of Observations." (For this essay Mr. Eisenhart was awarded the William Marshall Bullitt Prize in Mathematics, by Princeton University, in June 1933. EDITOR.) I have been working on various facets of the "translation" ever since, gaining, in the process, I believe, somewhat greater insight into the theory of measurement as a whole.

Before delving into fundamental concepts and principles of the theory of measurement as I see it, I wish to tell a story that has at least two messages for us here today: The story has to do with Coca-Cola vending machines on a particular Pacific island serving as a military base. These machines, unless empty, would automatically emit one bottle of Coca-Cola for each nickel inserted in the slot provided. But the costs had risen and bottles of Coca-Cola, according to the story, were now to be retailed at 6 cents a bottle. Unfortunately, the machine would accept only nickels. Experience revealed that the GI's would not put a penny in every instance into a little box that was placed there for the purpose. The military police did not have enough staff to have somebody posted there to see that the GI's did put a penny in the box. So, an operations analyst on the post was called in for advice. He had a statistical flavor in his background. He said that they should simply fill the machine up, but <u>on the average</u> put an empty bottle in every sixth place. In this way empty bottles will be arranged in the machine at random — the perfect strategy. If the empties were arranged in a systematic way, clever GI's would wait until somebody had bought an empty bottle and then they would charge the machine, and their enthusiasm added to his fury might be too much for the machine. By putting empties in at random, nobody could outguess the machine and all would be on an equal footing. The operations analyst certainly solved the seller's problem, in the sense that the seller was going to get his money — 6 cents per <u>full</u> bottle on the average — but he didn't have much of a heart for the purchasers on the other end of the deal. You may say that everything was all right, because everyone who used the machine would be fairly treated <u>in the long run</u> — provided he bought enough Coca-Colas — and by the strong law of large numbers the heavier the drinker, the more nearly certain that he would be getting cokes for 6 cents per drink. But the fellow on the island for just one day, with only one nickel to spend, he is either going to be lucky or unlucky. What about him? Is he treated fairly?

This story has two messages for us: First, correctness <u>on the average</u> does not guarantee satisfactory outcomes in <u>individual cases</u>. Thus, a single observation, or the average of only two or three, should not be used as if it were the average of a great many, without careful justification. Second, a consultant should consider his client's problem in its entirety, and help reach a full solution that is "best" from his client's viewpoint — he should not foist upon his client a clever solution to only a part of the problem, leaving the client to take the rap unprotected when the "worst" eventuality actually happens. Some treatments of the theory of errors fall, I fear, in this latter category.

Now, with these casual introductory remarks off my chest, I must buckle down to the serious business of my assigned topic —

What is measurement? Briefly stated, measurement is a process consisting of a sequence of steps or operations that yield as an end result a number that serves to represent the amount, the degree, extent, magnitude, or quantity of some property of a thing — a number that provides an answer to the question "how much?" for someone to use for a specific purpose. The purpose for which the answer is needed determines the method of measurement employed; that is, the sequence of operations by which the number is to be obtained; and also the precision and accuracy that are requisite.

When a magnitude is determined by the use of instruments whose indications yield directly the numerical value of the magnitude, the process is called direct measurement; and the result obtained, a direct measurement. Examples are: measurement of a length by a scale, of mass by a balance, of electrical resistance by a Wheatstone bridge, and of period of time by a clock. On the other hand, when determination of the magnitude(s) of one or more directly-measured quantities that bear a known relationship to the quantity under investigation, the process is called indirect measurement; and the result obtained, an indirect or derived measurement. If, for example, the volume of a spherical ball is computed from a direct measurement of its diameter, by means of the formula $V = \pi D^3/6$, the result is a derived measurement. If, on the other hand, the volume were determined by measuring directly, with a graduated vessel, the volume of liquid it displaces from a filled container, the result would be a direct measurement, even though arrived at by a roundabout procedure.

Today I shall limit my discussion, for convenience, to direct measurements, and direct-measurement processes.

A direct-measurement process is essentially a production process, the "product" being the numbers, that is, the measurements, it yields. There are two aspects of this process, the quantitative and the qualitative. The quantitative aspect consists of the readings or the observations themselves, which are the end product of the process and are in the form of numbers. The qualitative aspect consists of the manipulation of an instrument (or apparatus) and the taking of readings by someone, or by an automatic recording device, under prescribed conditions in accordance with specific instructions (i.e., rules of procedure). Thus, the factors that enter into the measurement of any quantity are the observer-apparatus combination employed (i.e., the person(s), the apparatus, and all the auxiliary materials, such as reagents, sources of illumination, etc.), the conditions under which the measurement operations are carried out, and the instructions followed.[1]

A characteristic of direct measurement is the disagreement of repeated measurements of allegedly the same quantity. Experience shows that repeated measurement of the same magnitude generally results in a series of non-identical numbers. To explain these discordances we introduce the concept of errors, which we interpret to be the manifestations of variations in the execution of the process of direct measurement resulting from "the imperfections of instruments, and of the organs of sense," and from the impossibility of achieving (or even specifying with a finite number of words) the ideal of perfect control of conditions and procedure.

It is, of course, highly desirable that our measurements be reliable, by which I mean not that they are totally free from error — this we can never achieve — but simply that such errors as they do contain are negligible in the sense that decisions or conclusions based upon the measurements as they stand will not differ in any important respect from the decisions or conclusions that would follow if the errors they contained could be and were removed.[2]

[1] For an excellent discussion of the qualitative aspects of measurement from an operational point of view, see W. A. Shewhart, Statistical Method from the Viewpoint of Quality Control (edited by W. Edwards Deming), The Graduate School, U. S. Department of Agriculture, Washington, 1939, p.130ff.

[2] Compare Shewhart, loc. cit., Rule 1 (p. 88) and Rule 2 (p. 92).

The degree of reliability required of a set of measurements depends primarily on the uses for which they are intended, but one should not ignore the requirements of other uses to which they are likely to be put. A set of measurements whose reliability is unknown is worthless; worse, it may be dangerous. A man is to be pitied who must of necessity reach a decision in some matter and to guide him has only data of inadequate or unknown reliability. In such a case he is forced to act much as did Steyning 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."

The reliability of a set of measurements as a basis for decision in some particular respect is, strictly speaking, unknowable, but can usually be inferred -- but not without some risk of being incorrect -- from the estimated precision and conjectured limits to the possible bias, that is, from the inferred accuracy, of the process by which the measurements were obtained. By the bias, or systematic error, of a direct-measurement process we mean the magnitude of its tendency to measure something other than what was intended; by its precision, the closeness together, that is, the degree of agreement amongst, repeated measurements of the same fixed quantity; and by its accuracy, the comprehensive term, the closeness of such measurements to the actual magnitude concerned. It is most unfortunate, I feel, that in popular parlance we often talk of "accuracy and precision," because accuracy includes "precision," but the converse is not necessarily true. It is less confusing, therefore, if we talk about mutually distinct concepts such as precision and bias (or systematic error, as it is often termed, with less stigma, perhaps). Indeed, if I succeed today in accomplishing no more than making clear to you the distinction between these three terms — precision, bias, and accuracy — our time together, I feel, will have been well spent. I hope, however, to accomplish a bit more if time permits.

The distinction between accuracy and precision as applied to measurement, measurement processes, and measuring instruments, is as follows:

(1) The accuracy of a measurement process pertains to degree of conformity to the truth of measurements generated by repeated applications of the process under fixed circumstances.

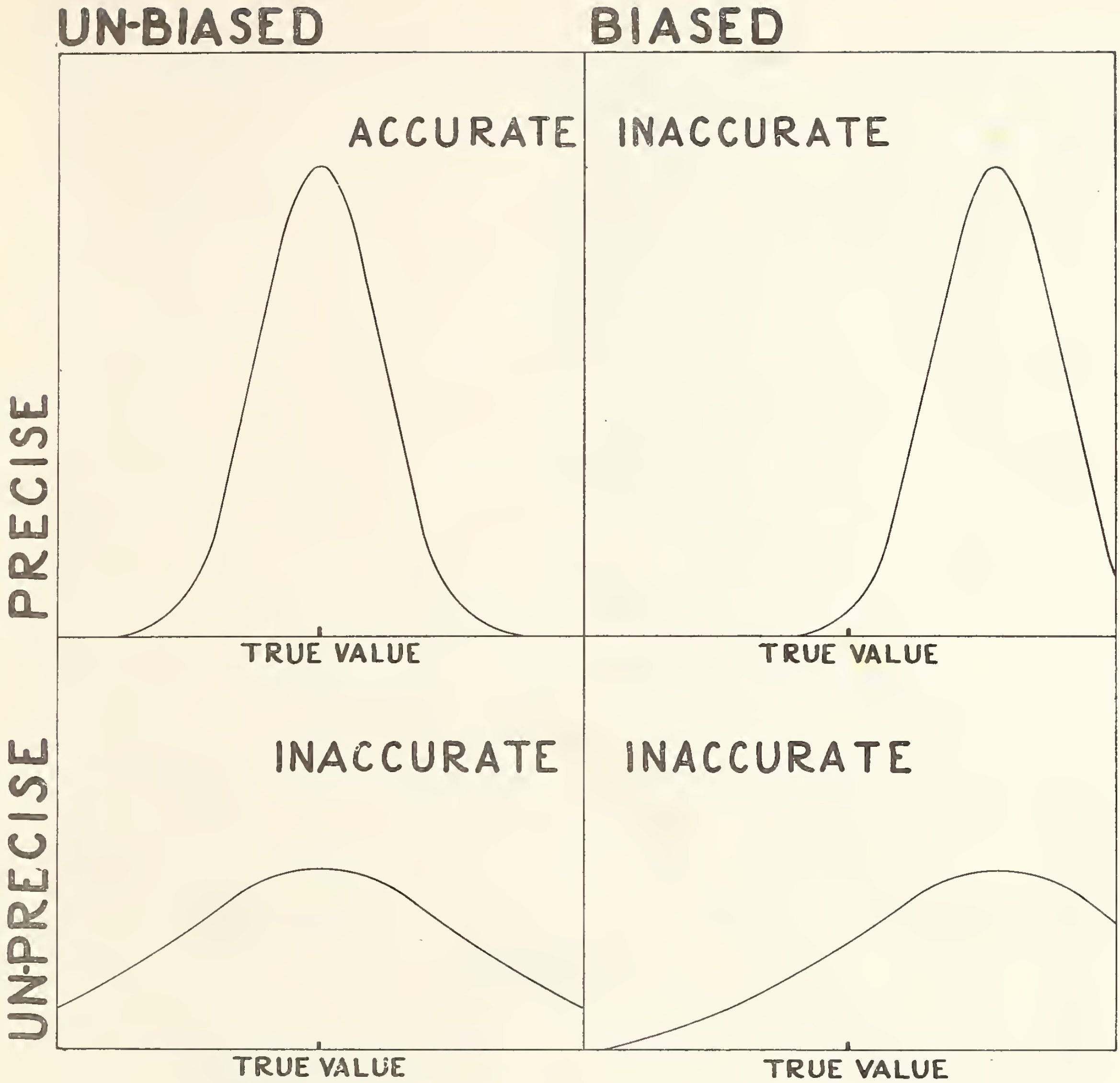
(2) The precision of a measurement process pertains solely to the degree of conformity of the measurements among themselves; and hence to the degree of their conformity to the average value characteristic of the process in the particular circumstances concerned, quite irrespective of whether this average value is or is not the 'true value'.

In other words, accuracy refers to the closeness of the measurements to the 'true value' — closeness to some reference or standard value accepted as the truth — whereas precision refers merely to their closeness together. Thus, accuracy expresses a relation to a value external to the measurement process; precision, to a value internal to the process.

An accurate method of measuring some quantity is, therefore, a method that is both precise and unbiased, in the sense that it yields measurements that are closely clustered and centered on the 'true value'. (Such a situation is portrayed in the upper left-hand quadrant of Figure 1.) If the measurements are closely clustered, but centered on some value other than the 'true value', then the method is precise, but biased, and hence inaccurate. (The upper right-hand quadrant of Figure 1 portrays such a situation.) If the measurements are widely scattered, but nevertheless are centered on the 'true value', the method is unbiased, but imprecise, and hence inaccurate. (This situation is depicted in the lower left-hand corner of Figure 1.) Finally, if the method is both biased and imprecise, it is a fortiori inaccurate. (The lower right-hand quadrant of Figure 1 illustrates this situation.)[3]

From what I have just said, however, I most certainly do not want you to infer that an unbiased procedure is always to be preferred to a biased one. Indeed, a procedure with a small bias and a high precision can be more accurate than an

[3] You will, I believe, find that the foregoing distinctions are consistent with those drawn between accurate and precise in the synonymic article that appears under "correct" in Webster's New International Dictionary.



INTER-RELATIONS
OF
BIAS, PRECISION & ACCURACY

FIGURE 1.

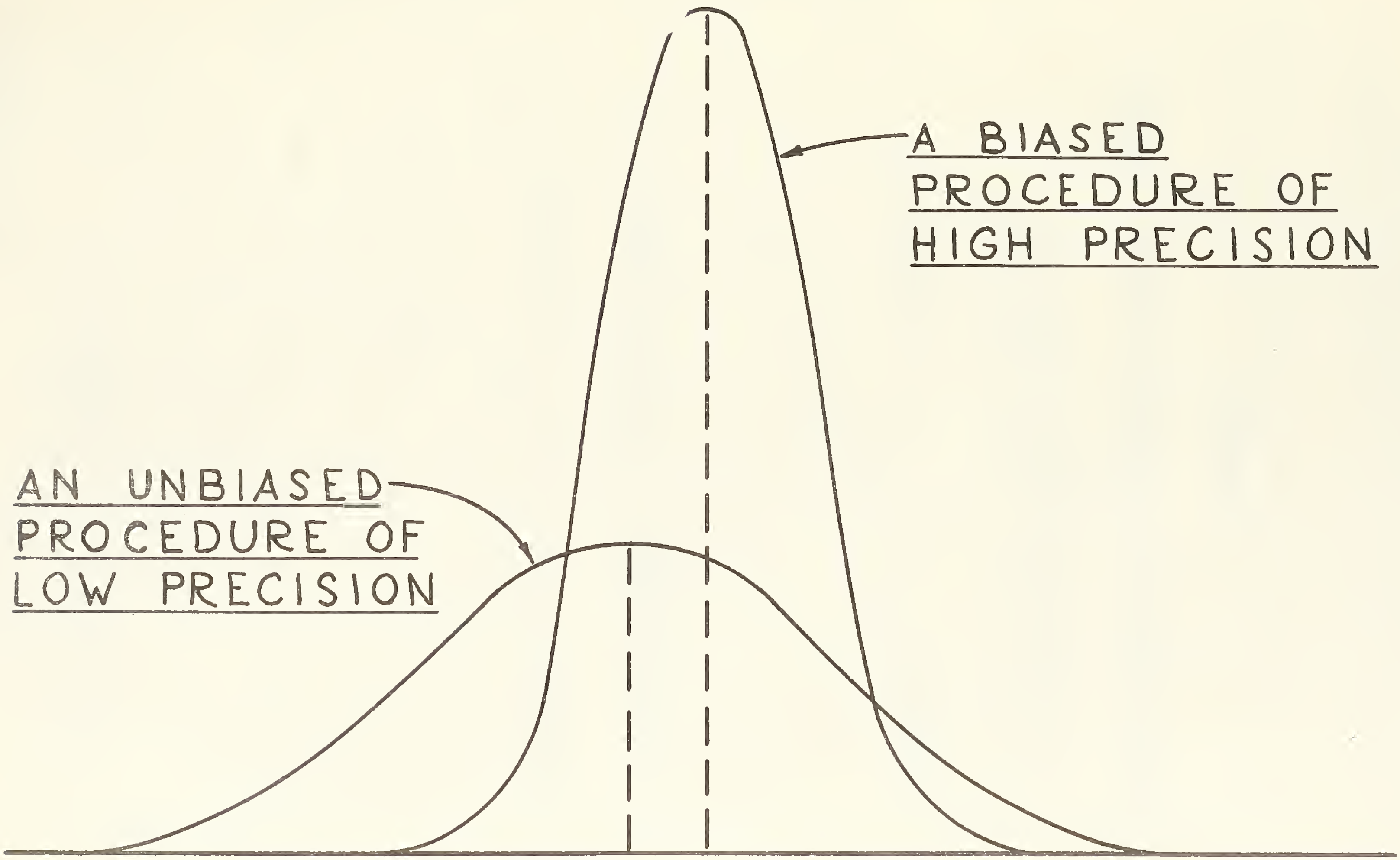


FIGURE 2 - COMPARISON OF AN ACCURATE BIASED PROCEDURE WITH AN INACCURATE UNBIASED PROCEDURE

unbiased procedure of low precision. (See Figure 2.) 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 be strictly correct on the average. Consider carpentry: I sincerely doubt whether even the best of carpenters hit nails with absolutely no bias (up, down, right, or left) on the average; but good carpenters surely don't miss the nail altogether very often; and are certainly to be preferred to an imprecise but well balanced novice who hits most every spot within six inches of the nail, with absolutely no bias in the long run, but rarely if ever hits the nail itself. This we must remember: in practical life we rarely make a very large number of decisions of a given type -- we can't wait to be right on the average -- our decisions 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.

It is clear from what I have said earlier that the problem of bias, or systematic error, would be licked if we could be sure that a particular direct-measurement process measured exactly what was intended. This goal is in effect achieved in the writing of performance specifications for materials and products by including within the performance specification itself a detailed specification of how a particular magnitude is to be measured, or by referencing a specific method of measurement given in some supplementary document, and then accepting the method of measurement so prescribed as defining the "true value" of the quantity concerned for the purposes of the performance specification itself. Thus, in the Specifications for Government Synthetic Rubbers, the tolerances stated for the "viscosities" of the several synthetic rubbers relate to "viscosity" as defined by the method of measurement spelled out in detail elsewhere in these Specifications; and, in the new Federal Specification for "Rubber; Cellular, Latex Foam," it is stated that "latex foam rubber shall show a compression set not greater than 20 percent when tested as described in [Section] 433," which is entitled "Compression set" and states that "the compression set shall be determined as described in method 11005" of Federal Specification ZZ-R-601: Rubber Goods; General Specifications (Methods of Physical Tests and Chemical Analyses).

This operational approach to the definition of the true values of physical and chemical quantities brings us, however, face to face with another fundamental question: in what sense can we say that a particular direct-measurement process defines an unique magnitude, the value of the quantity so determined, when experience shows that repeated application of the process under fixed circumstances yields a sequence of non-identical numbers? What is the value thus defined?

The answer takes the form of a postulate about direct-measurement processes that has been expressed by N. Ernest Dorsey (on p. 4 of his "Velocity of Light," Transactions of the American Philosophical Society, 34, 1-109, October 1944) as follows:

> "The mean of a family of measurements — of a number of measurements of 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."

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 Direct Measurement. A mathematical justification for it can be found in the Strong Law of Large Numbers, a theorem in the mathematical theory of probability discovered during the present century.

Furthermore, consideration of the conditions under which the Strong Law is valid furnishes an indication of the circumstances under which the Postulate of Direct Measurement is likely to be effective in practice.

It will suffice here today for us to note that the sole aim of the Postulate of Direct Measurement is axiomatic acceptance of the existence of a limit approached by the arithmetic mean of a finite number $\underline{n}$ of measurements generated by any direct-measurement process as $n \to \infty$; and it should be noted that it says nothing on 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 non-identical 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 or observations "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 <u>the</u> value of the magnitude that his first set of measurements or observations were striving to express.

The foregoing can be expressed mathematically as follows: on some particular occasion, say the "i^{th}", we may take a number of successive measurements by a given direct-measurement process under certain specified circumstances. Let

(1) $x_{i1}, x_{i2}, \ldots$

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 under the same fixed set of circumstances. Each such fresh "start" would correspond to a different value of "i." If, as we shall assume, the measurement process concerned when applied under these circumstances obeys the Strong Law of Large Numbers, i.e., if the Postulate of Direct Measurement is applicable, it follows that we may expect the sequence of cumulative arithmetic means on the i^{th} occasion, namely,

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

to converge to μ, a number that constitutes the limiting mean associated with this direct-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 (see, for example, William Feller, <u>An Introduction to Probability Theory and its Applications</u>, vol. <u>1</u>, New York: John Wiley & Sons, Inc., 1950, p. 207) does <u>not</u> 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,...), the instances of non-convergence to μ will be rare exceptions. In other words, 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 Direct 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 as the mean or "center of gravity" of the infinite conceptual population of all measurements x that might conceivably be generated by the direct-measurement process concerned under the specified circumstances.

With this as background, we are now in a position to consider the mathematical definition of the precision of a direct-measurement process under a fixed set of circumstances. By definition, the precision of the process has to do with the closeness together of the individual measurements generated by the process under these fixed conditions. Otherwise expressed, it has to do with the closeness together of the two individual measurements constituting an arbitrary pair. Let us assume for the moment that under the circumstances chosen the direct-measurement process gives rise to a sequence (1) of completely <u>homogeneous</u> measurements — the full meaning and import of this qualification will become apparent as we proceed. Let us now consider the individual measurements (1) to be grouped <u>arbitrarily</u> into <u>pairs</u> giving rise to the derived sequence of <u>differences</u>

(3) $\quad d_1, d_2, \ldots, d_n, \ldots$

where the additional subscript i has been omitted for convenience. Some of these differences will be positive, and others negative, and it is not difficult to show that whatever be the (finite) value of the limiting mean μ associated with the sequence (1), the limiting mean δ associated with the sequence of d's, (3), will be identically 0. Consequently, the limiting mean of these differences is utterly useless as a measure of compactness of the original sequence (1). On the other hand, it is clear, I believe, that just as each individual measurement x is striving to express the value of the limiting mean μ, so also is each of the differences d, if its sign be neglected, striving to express the characteristic spread between two arbitrary measurements x.

To get rid of the signs of the d's, let us therefore, consider instead of (3) the sequence

(4) $\quad d_1^2, d_2^2, \ldots, d_n^2, \ldots$

of the <u>squares</u> of these <u>differences</u> between <u>arbitrary pairs</u>. It is not difficult to determine conditions on the measurements themselves, that is the x's, sufficient to ensure that the sequence (4) of the d^2's values will also obey the Strong Law of Large Numbers, and be associated with a limiting mean, say $\Delta^2 = 2\sigma^2$, where $\underline{\sigma}$, termed the <u>standard deviation</u> of the measurements themselves (the x's), is simply the <u>radius of gyration</u> of the aforementioned infinite population of x-values about its center of gravity μ. In other words, σ^2 is simply the average value of $(x-\mu)^2$ in this infinite conceptual population of possible measurements x.

Since the <u>precision</u> of the process obviously <u>decreases</u> as the value of $\underline{\sigma}$ increases, and vice versa, it is natural to take some inverse function of $\underline{\sigma}$ as a measure of <u>precision</u>. Thus, Gauss (c.f. Gauss, <u>Theoria Motus Corporum Coelestium In Sectionibus Conicis Solem Ambientium</u>, Hamburg, 1809, Article 178) adopted as his <u>modulus of precision</u> the quantity $h = 1/\sigma\sqrt{2}$, which we see to be the square root of the reciprocal of the limiting mean of the d^2 sequence (4).

Mathematically the foregoing discussion can be carried out equally well in terms of the absolute (un-signed) values of the d's, instead of in terms of their squares. Such an approach, however, has several disadvantages. In the first place, the limiting mean, say δ', with a sequence analogous to (3) but in which the signs of the d's are ignored, has a less vivid <u>geometrical</u> - or should I say <u>mechanical</u> - interpretation than has σ; and δ' works out to be equal to $a\sigma$, where the constant $\underline{a}$ depends on the particular functional form of the distribution of the measurements x about their limiting mean μ. Secondly, as we shall now see, <u>components of error</u> are additive in terms of squared quantities such as σ^2, so that in this sense σ^2 is a more appropriate measure of the dispersion of the x's about their limiting mean μ than is σ itself or any constant multiple of it.

In the foregoing we assumed for convenience that the individual measurements forming the sequence (1) were completely <u>homogeneous</u>. In practice this is rarely the case and a more common situation is that in which a sequence (1) consists of a series of "sections" with the measurements in any one section being homogeneous with respect to each other, and pairwise more close together on the average than two measurements one of which comes from one section and the other from another. In the simplest of such cases, if we form a sequence such as (4) composed of the squares of differences between <u>arbitrary pairs</u> of measurements <u>from within each of the respective sections</u>, then the limiting mean of such a sequence of d^2's will be of the form $2\sigma_w^2$, where σ_w^2 is the <u>within-group variance</u>. If, on the other hand, we form arbitrary pairs consisting of <u>one measurement from each of two different sections</u>, then the limiting mean of a sequence of such d^2's will be of the form $2(\sigma_w^2+\sigma_b^2)$ where σ_b^2 is the <u>between-group variance</u>.

In such a situation, if $\bar{\bar{x}}_n$ is in fact the average of a total of n = km measurements, composed of m measurements from each of k different sections, then over a (infinitely) large number of such experiments, i.e., different "starts," the average value of $(\bar{x}_n - \mu)^2$ will be

$$(5) \qquad \frac{1}{k}\left(\sigma_b^2 + \frac{\sigma_w^2}{m}\right) \quad ,$$

from which it is clear that, if σ_b^2 is at all sizeable compared to σ_w^2, then $\bar{x}_n$, for fixed $n = km$, will have greater precision if based on a large number k of different sections, with only a small number m of values from each section.

The foregoing can be interpreted, I believe, in the language of your science, as I understand it, somewhat as follows: Let "sections" correspond to different prints from a photographic negative so that measurements within a particular "section" are repeated determinations of the distance between two points on the ground, say, as determined through successive measurements of this distance on this single print. Let us now suppose further that there are two points on the first, the distance between which on the ground is accurately known, having been determined to everybody's satisfaction with sufficient accuracy by the man on the ground with the invar tape and other auxiliary apparatus of a professional surveyor. Let the true value of this distance be denoted by X and let us suppose that the problem is to determine the distance between two other points, the true value of which is, say, Y. One method of doing this would be to make successive independent measurements x_1, x_2, x_3, ..., x_m, of the distance between the two standard points on the photographic print, and an equal number of independent measurements $y_1, y_2, \ldots, y_m$ of the distance between the other two points on this same photographic print. One could then take as an estimate of Y the quantity

(6) $$Y = X + (\bar{y}-\bar{x}) ,$$

where $\bar{y}$ and $\bar{x}$ are the averages of the y and x determinations respectively, which we assume are all mutually independent and have precision implied by individual variances of σ_w^2. The variance of Y as an estimator of Y will then be $2\sigma_w^2/m$. However, Y may be a biased estimator of Y, the magnitude of the bias, β_1, being a property of this first print. One could check this latter by calculating a Y_i from each of k different plates $(i=1,2,\ldots,k)$ and then checking to see whether the quantity

(7) $$\sum_{i=1}^{k} (Y_i-\bar{Y})^2/(k-1) ,$$

where $\bar{Y}$ is the average of the Y_i, is significantly larger than $2\sigma_w^2/m$. If it is significantly too large — and we have statistical tests for answering this question — then we must conclude that (6) is not an estimator of $\frac{2\sigma_w^2}{m}$ of $\sigma_b^2 + \frac{2\sigma_w^2}{m}$, where σ_b^2 denotes the variance of the biases $\beta_1, \beta_2, \ldots$ about their limiting mean β^*, say. The bias of the photoprint method is measured by β^*. In such a case the variance of the over-all average $\bar{Y}$ obtained from all prints will be given by

(8) $$\frac{1}{k}(\sigma_b^2 + \frac{2\sigma_w^2}{m}) \quad ,$$

where the additional "2" — in comparison to (5) — comes from the fact that on each print the comparison of measurements on two different distances is involved. Clearly, whether it is desirable to take a large number of measurements on only a few prints, or only a few measurements on each of a large number of prints, will depend on the relative magnitudes of σ_b^2 and σ_w^2 . Furthermore, it is evident that instead of considering different prints from the same photographic plate, one might consider instead different photographic plates obtained on different flights over the region by an airplane, and so forth and so on.

In applied science one often speaks of "repeating the determination" of some quantity. In a setting such as the foregoing one should be clear on exactly what one means by a "repetition." Does it mean more measurements of the same kind on the same print by the same observer-equipment-procedure combination, or would the same observer-equipment-procedure combination be employed using different prints; or, would various but equivalently-trained observers be employed with the same or similar equipment, using the same or similar procedures, at the same or various places, using the same or different prints, from the same or different negatives; and so forth? Clearly it is not possible to talk ambiguously about the <u>precision</u> of a particular method of measurement without indicating the character of the "repetitions" involved in generating the sequence of like measurements with respect to which the "precision" is supposed to apply.

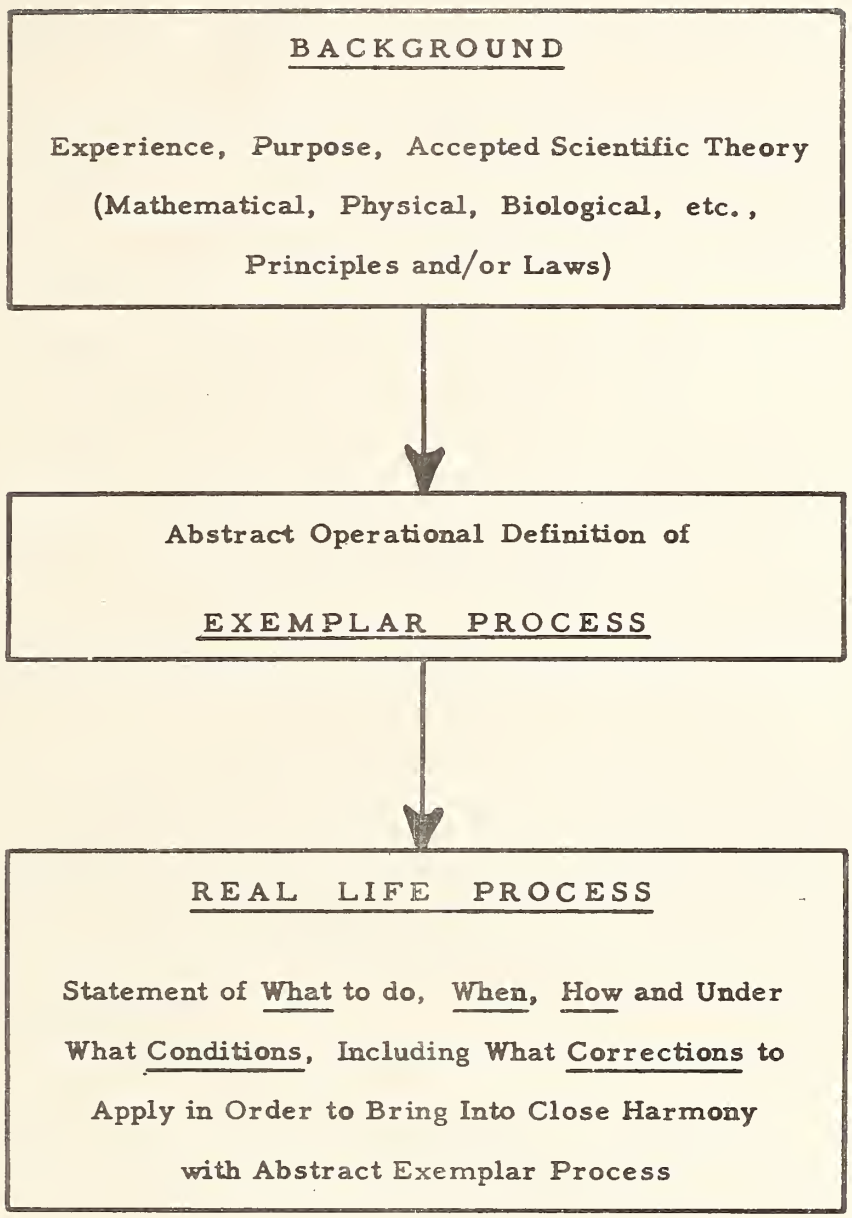
There is also the problem of how to proceed in making "repeated measurements" so that the results obtained will be <u>independent</u> in the statistical sense; if one is measuring the distance between two points on a print with the same calibrated scale over and over again, it is exceedingly difficult, if not impossible, to obtain <u>independent</u> readings unless one deliberately introduces a random positioning of the scale in each instance. Alternatively, we might use a series of different graduated scales of the same general type but which had different calibration corrections. In this way one might help the "rounding errors" and "reading errors" to balance out. The use of measuring rods with unevenly spaced scale divisions for this purpose was discussed by P. C. Mahalanobis, F.R.S., in a lecture at the National Bureau of Standards, on November 13, 1946; a summary of his lecture will be found in the ASTM Bulletin for January 1947, pp. 64-66. I commend this matter to your attention.

Finally, I feel that a few words are in order on the subject of "true value." Earlier in discussing the <u>bias</u> of a direct-measurement process I remarked that the <u>bias</u> is defined to be the difference, say, $\mu - \tau$, between the value μ that the process measures — its limiting mean — and the true value, τ. This immediately raises the question: How is the "true value" of a property or characteristic defined?

To answer this question we begin first by noting with P. W. Bridgman that a property or characteristic of the physical world is defined, in the last analysis, by specification of a method of measuring its quantity:

> "What do we mean by the length of an object? We evidently know what we mean by length if we can tell what the length of any and every object is, and for the physicist nothing more is required. To find the length of an object, we have to perform certain physical operations. The concept of length is therefore fixed when the operations by which length is measured are fixed; that is, the concept of length involves as much as and nothing more than a set of operations by which length is determined. In general, we mean by any concept nothing more than a set of operations; the concept is synonymous with the corresponding set of operations. If the concept is physical, as of length, the operations are actual physical operations, namely, those by which length is measured; or if the concept is mental, as of mathematical continuity, the operations are mental operations, namely those by which we determine whether a given aggregate of magnitude is continuous... We must demand that the set of operations equivalent to any concept be a unique set, for otherwise there are possibilities of ambiguity in practical applications which we cannot admit."
>
> (P. W. Bridgman, The Logic of Modern Physics, Macmillan, New York, 1927, pp. 5 and 6.)

It should be clear to us from what Bridgman has said that if all of you and I are to agree on what is meant by the length of this blackboard along its lower edge, then we must first come to an agreement on a sequence of operations that is to be taken as defining the concept of "length" in this case. This done, the true value of the length of this blackboard along its lower edge will then be uniquely determined by the limiting



THE EVOLUTION OF A REAL-LIFE DIRECT-MEASUREMENT PROCESS

FIGURE 3.

mean associated with the agreed-upon procedure. In brief, as W. Edwards Deming has observed, the true value of a characteristic is in fact defined by agreement among experts on an EXEMPLAR PROCESS for measurement of the characteristic. (W. Edwards Deming, Some Theory of Sampling, John Wiley & Sons, New York, 1950, pp. 15-17.)

Dr. Deming actually uses the term "preferred procedure" and notes 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;" and "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. The precision of a random or stable procedure, however, may be measured and known."

As I see it, the evolution of a real-life direct-measurement process is essentially as shown in Figure 3. From experience we are aware of recognizable changes in things going on about us, and we say that these take place with the passage of time — we have a "feel" for what we mean by "time;" we can talk about it with one another, etc.; but we find ourselves beset with many difficulties when we try to define exactly what we mean by "time." If we try very hard to define "time," we will find that we wind up by agreeing that we accept the successive occurrences of certain recognizable events as "measuring" the passage of "time." For example, we may note the accumulation of sand in the bottom chamber of an hourglass, or we may count the oscillations of a pendulum, as in a grandfather's clock, or the oscillations of an atomic system as in an "atomic clock." Similarly, with regard to <u>length.</u> We are aware of the experience of the separation of objects in space. From our study of Euclid we feel that we have "knowledge" of the meaning of the "length" of a line, that is, of the distance between two points; but remember that Euclid was writing abstract mathematics and not physics. If we ask so simple a question as what is the "length" of this blackboard along its bottom edge, we find that all seems to be very clear until we begin to get pickayune about it. If our knowledge of physics goes only so far as awareness of molecules, we may say that the "outside point" at the left end is by definition the outsidemost point on the leftmost molecule. Then someone rushes into the room and tells us about the Mendelian theory of inheritance. We listen attentively, turn this new idea over in our minds, and say to ourselves, "Very interesting, but it does not seem to have any bearing on how we ought to define length." Then, we are told, perhaps about atomic theory, and this definitely worries us, for we see that we will now have to define the endpoint on the left as the center, say, of the leftmost electron of the outermost atom — but this brings up new difficulties, for the electron is whirling around and not staying in any one place, so we will have to take the outermost point of its path. But how are we going to see it, because if we shine some light over here, the light "waves" or "particles," according to what you are believing this morning, will push the electrons out of their normal courses and so when we try to see the outside point we move it. Clearly this is getting us nowhere fast!

We therefore pull ourselves up to a halt with a reminder that we wanted to know the "length" of the blackboard along the bottom edge there for a purpose: so that we could cut off a board of the right "length" and nail it on there to form a chalk tray. We then call on experience again. We have seen yardsticks or metersticks lying around, and so from these we abstract the concept of a rigid bar ideally marked out with graduations that are perfectly equidistant apart, the space between two such graduations being one unit of "length." Fine! All we have to do now is count how many units of length there are in the bottom edge of this blackboard. So we try it. We find that it is between 205 and 206 units, say. A unit is rather large, so that a half or a quarter unit would look unsightly sticking out endwise on the blackboard, so we subdivide our "unit," or use some other ingenious means of getting a finer distinction. The purpose at hand determines how far we attempt to proceed with this business of making finer and finer distinctions. If our bar has been lying on the radiator so that each of the "units" on it has grown a little and is really bigger than a real unit, and the lumber is outdoors in the cold, so that our bar will shrink and have "units" that are really too small when we take it outdoors to measure the lumber, then we apply temperature corrections based on physical theory. And so forth and so on. I believe you get the idea. I can hardly make it more precise than this in the time I have here.

The important thing is to let the purpose determine the refinement that you are willing to go to. And one should keep in mind not only the immediate purpose but also other uses to which the result might be put.

For example, I am told that the makers of topographic maps from on-the-ground surveys generally draw in the contour lines by intuition and the general "feel" of the landscape as they stand there looking at it. The resulting contours may be adequate for some purposes, but not for all. The following story has always amused me in this connection: The elevation of Shongum Lake in New Jersey was determined a good many years ago by C. C. Vermeule to be 698 feet. On the other side of the mountain east of the lake, he sketched in a ravine using uniformly interpolated contours between two roads along which levels were run, and which the ravine cuts across. Years later a reservoir was built in this ravine to supply the insane asylum near Morris Plains, New Jersey; and in subsequent maps the reservoir was sketched in. This reservoir, from the adjacent contours on such maps, had an apparent elevation of 640 feet. In consequence someone got the idea of building a gravity feeder line from Shongum Lake, via a nearby saddle point, to the reservoir, so that the Lake could be used to supplement the reservoir. Unfortunately, a special survey revealed the actual height of the reservoir to be about 720 feet - quite a bit higher than the 640 feet suggested by the contours - so that the water from the reservoir would have, in fact, flowed into the Lake! Photogrammetry, I am told, provides a means of drawing in numerous contours quite accurately, and at much less cost than by ground surveys. Let's hope that you succeed in making them accurate enough for all uses likely to be made of them.

SUPPLEMENTARY DISCUSSION

(Following Mr. Bicking's Paper)

MR. KATZ:

.....

I wonder if Captain Reading would like to say something about this whole business. I referred earlier to the fact that, by and large, European photogrammetrists are very much concerned with statistical method, and the theory of errors. On numerous occasions they have pointed out that, with always present exceptions, we over here have not yet seen fit to apply these things to American photogrammetry.

CAPTAIN READING: That is quite correct. I think we in photogrammetry in the United States have not had time to study the theory of measurement. We have been mystified a bit by the second diagram of the four drawn by Dr. Eisenhart. Dr. Eisenhart seemed to prefer the narrow range with bias to a truer average with wider dispersed values.

There has been a tremendous amount of publication and discussion in the past that has failed to register or strike home — largely because of differences in terminology and lack of agreed-upon definitions for these terms. Clarification of our terms would get us talking the same language. I remember in 1948 the Europeans seemed to set great store on trying to run down the difference between systematic and accidental errors. The theory seemed to be that there was some distinction that should be made between the two and then we had the process nicely corrected and were doing the most efficient thing. But the reason we have been skeptical about applying all these mathematical and statistical methods is because we were always stopped by this type of question: should we spend our money on first order theodolite observations that are read to one second, or should we spend our money running photograph after photograph through a coordinate setting machine and repeating flight after flight in the air when we can read say to only 12 seconds. Now most of us have said that we would use the theodolite and forget about the trips through the plotting machine and through the air.

But there is a very real need to know just exactly where to draw the line, because sometimes we are up against the situation where one approach is very expensive and the other much less so.

I would like to ask Dr. Eisenhart to clarify this preference for the narrowed spread with bias over the unbiased wider spread. What are the criteria you use to know which is better?

DR. EISENHART: I'd certainly better clear this up. The comparison and choice that I made is really only available when you have some very fine measurement procedure — an exemplar procedure — that all will agree yields — at least in the limit — the "true value" of the magnitudes concerned. You see in Figure 2 there is a mark on the horizontal axis to indicate the "true value" and you have to know where that value is in order to be able to decide whether a given distribution of readings is centered on that point, or near to it, or far from it. For example, if you are working on certain kinds of ground surveys, there are very likely some types of distances on the ground that all will agree that the fellow on the ground with an invar tape can measure more accurately than a fellow can from an aerial photograph made from 40,000 feet. If you have such a method of measurement, which may be very expensive but which professionals in the field agree upon as giving you what we will call the "true value," then this can serve as our reference system. Suppose now that we have two alternative inexpensive methods where one of these tends to yield more widely dispersed values but without bias — that is individual measurements obtained by this procedure are widely dispersed but as a group are centered upon the true value. This situation is represented diagrammatically by the flatter curve in Figure 2. With this method if you took, say, 100 readings, the average would be almost certain to be very close to the true value; but a single reading, or even an average of say five readings, would have a good chance of being quite far off. The other of these inexpensive methods, we assume, is characterized by the taller of the two curves in Figure 2, that is, individual readings by this method are hardly dispersed at all, but they have a little bias. Now it is clear from the figure, I believe, that if cost or time or some other restriction limited us to a single reading, then we would have a much better chance of getting a reading close to the true value by using this second method than we would have with the first. On the other hand, since this second more precise procedure has a small positive bias, the average of a large number of readings by this method will be almost certain to be too high. Now here is the important point: if the cost per reading is the same for either of these two procedures, and if neither cost nor time place a severe limit on the number of readings that you may take, then there will be a number of readings n_0 such that for averages of n_0 or more readings the probability of such an average being within $\pm\varepsilon$ of the true value, for any assigned $\varepsilon > 0$, will be greater, i.e. closer to 1, for the first method — the one with more widely dispersed individual values — than for the second; indeed, as the number of observations averaged together tends to infinity, this probability will tend to 1 for the first method, and to 0 for the second method for all values of ε less than the magnitude of the bias of the second method. On the other hand, as we have already seen, for only single readings, or averages of small numbers (much) less than n_0, the second method may put answers closer to the bull's-eye. If there is a difference in cost per reading with the two methods, then the same principles apply, although the computation may be slightly more difficult: if, say, one can take ten readings with the widely-dispersed unbiased method for the same cost as a single reading by the narrow biased method, then it may, perhaps, be the case that averages of five readings by the narrowly-dispersed biased system would have a greater probability of being closer to the true value than would averages of 50 readings by the widely-dispersed unbiased method; but the situation may be reversed in the case of averages of 500 readings by the unbiased method with the wide dispersion and 50 readings by the biased system of narrow dispersion. I hope that I have succeeded in making clear to you that there are certainly circumstances where one might do well to choose the second biased method in consequence of its small bias in relation to its high precision, even though it will tend to shoot uphill a bit.

I don't want you to focus too much of your attention on this question of biased versus unbiased, for bias or absence of bias is only important insofar as it affects accuracy — the important thing is accuracy! A method of measurement that yields accurate determinations of a quantity α also provides a means of obtaining accurate determinations of functions of α. Thus, if $\beta = f(\alpha)$, and a is an accurate determination of α, then $b = f(a)$, the derived value, will likewise be an accurate determination of β; but, in general, if a is an unbiased determination of α, b will be a biased determination of β. It may be possible to derive from a an unbiased determination of β, say b', but b' may have somewhat less precision than b and possibly, as a consequence, less accuracy.

For example, suppose that the problem is to determine the area of a circle and that this is to be done by measuring its diameter. Suppose that the true area is $\frac{\pi D^2}{4}$. Now suppose further than your diameter-measurement method is <u>unbiased</u>, and individual diameter measurements <u>d</u> are normally distributed about D with variance of σ^2. Then, given <u>n</u> independent diameter measurements $d_1, d_2, \ldots, d_n$, the average of these <u>n</u> measurements, $\bar{d}$, will be an unbiased estimator of D, with variance $\frac{\sigma^2}{n}$. The square of $\bar{d}$, however, will be an unbiased determination <u>not</u> of D^2, <u>but</u> of $D^2 + \frac{\sigma^2}{n}$; and be distributed about this latter magnitude with variance $\frac{4D^2\sigma^2}{n} + \frac{2\sigma^4}{n^2}$, under the assumption that the d's are normally distributed about D.

Since

$$s^2 = \sum_{i=1}^{n} (d-\bar{d})^2/(n-1)$$

is an unbiased estimator of σ^2, one could obtain an unbiased determination of D^2 by subtracting $\frac{s^2}{n}$ from the square of $\bar{d}$, but the price of this adjustment would be to add to the foregoing expression for the variance of the square of $\bar{d}$, an additional term $\frac{2\sigma^4}{n^2(n-1)}$. Since <u>n</u> or some higher power of it appears in the denominator of all of the terms of the foregoing expressions except the term in D^2 alone in the first expression, it is clear that these problems of bias and corrections for it (with a consequent inflation of the variance) all become unimportant if we use a large enough number of observations _n_ in the first place. Furthermore, even for small values of _n_ we might not wish to apply the foregoing correction to the square of $\bar{d}$ for the following reason: If, as we have assumed, the distribution of individual diameter determinations _d_ is symmetrical with respect to the true diameter D — we have actually assumed that the distribution of _d_ is normal about D — values of $\bar{d}$ greater than D and values of $\bar{d}$ less than D will occur equally often in practice in the long run. Also, for any sensible combination of _σ_ and _n_, negative values of _d_ ought not to occur in practice, so that the square of $\bar{d}$ will exceed D^2 as often as it is less than D^2 in the long run. In consequence, we may say the $\pi(\bar{d})^2/4$ is a probability-wise unbiased estimator of the true area.

CAPTAIN READING: I wonder if you would clarify the distinction between the types of errors that you mentioned — presumably when you had some means of measuring, of determining the bias, you have what we call a systematic error, and you can eliminate it if you have some means of measuring the factor that put that error in there. Observations also contain what the Europeans call accidental errors, which they try to distinguish and separate out. I think that there is here some distortion of terms. I wonder if you'd care to clarify the definition of them. It seems to me that you left out something about accidental errors. I think that the terms random errors and gross errors instead of accidental errors might clarify the discussion a lot. Would you mind commenting on this?

DR. EISENHART: I used the term "random errors" to denote what are often termed "accidental errors" because I feel that the term "random" serves better than does "accidental" to signify the basic concept involved. It is my understanding that "gross errors" is the term used to signify deviations caused by actual blunders or mistakes of the observer, or by departures from the ideal operation of the process of measurement resulting from inattention of the observer to explicit or implicit requirements of the instructions. They are primarily "observer errors." "Gross errors" that are discovered tend to be large in magnitude — hence the term.

"Systematic errors" differ in my understanding from "gross errors" through their tendency to persist through several successive measurements, or perhaps, through an entire series of measurements, whereas the more usual types of gross errors are sporadic in occurrence and affect only a single observation or only a small number of successive measurements. The term "systematic errors" appears to be the general term for errors that manifest themselves as <u>shifts</u> in the central position of the readings, as <u>trends</u> in the readings, or as <u>oscillatory variations</u> in the position of the readings over long or short periods. Systematic errors that persist throughout an entire series of readings are more properly termed "constant errors."

A "constant error" may affect just one particular series of measurements by a given direct-measurement process, arising through some fault in the execution of the instructions on that occasion. If, on the other hand, the "constant error" is in fact "an error of method," resulting in the displacement of the limiting mean μ for that method even when properly applied, from the true value τ, then it is in effect a source of <u>bias</u>. The existence of constant errors can at least in principle be detected by taking several series of measurements at quite different times on standard material; but <u>bias</u> which is a characteristic of the method of measurement itself cannot be detected except by comparison of results obtained by different methods of measurement that are assumed (when appropriate corrections have been applied) to measure the same quantity.

Finally, experience shows that in the absence of gross errors and when all available sources of systematic error have been removed, or their effects eliminated by appropriate corrections, a sequence of measurements nevertheless exhibits fluctuations that may be considered to be the manifestation of the inherent vicissitudes of many minor uncontrolled factors; that the errors thus generated are unpredictable; and that their causes defy diagnosis and eradication. These "rock-bottom" variations behave like a series of random drawings out of a formal mathematical "urn" such as are considered in the theory of probability, and are in consequence termed "random errors." It is only to these random errors that the notion of a "law of error" strictly applies. Furthermore, <u>precision</u> has to do with the effect of these random errors. The <u>probable</u>

error or standard error of an average of the number of readings serves only to indicate the character of the uncertainty in this average resulting solely from the effects of the random errors.

It is my opinion that in some quarters there is a tendency to incorporate in the computation of probable errors or standard errors, components of error that are estimated magnitudes of the constant error or bias present; and that these should be reported separately and not built into the probable error. Consider the following example: In a particular direct-measurement process the limiting mean μ is sensitive to the actual diameter of a certain wire where it passes through a hole in the apparatus. The experimenter, realizing that the limiting mean depended upon the diameter of the wire, but not realizing how sensitive it was to this quantity, applied the necessary correction to the average of his measurements, using the nominal diameter of the wire as given on the spool. Some time later he discovered that variations in the diameter of the wire of the size such as occur along the wire from a single spool are sufficient in magnitude to make an important effect on the result. Unfortunately, he no longer had at hand the piece of wire used in the earlier experiment. He, therefore, measured the diameter of the wire from his spool at a large number of points along its length, and determined upper and lower limits which he believed would bracket the true diameter of the wire that he actually used. Now it seems to me that in such a case he should simply compute and report bounds to the possible bias of the end results of the first experiment that may have resulted from his failure to make this more refined diameter correction. The actual bias might have been zero — since he may have been lucky and have used a piece of wire that actually had the nominal diameter. He will never know what his actual bias was, but he can set limits on it. Now I feel that these limits on the possible bias should be recorded as such, and no attempt made to build the added uncertainty into the probable error of the mean that he reported earlier.

If, on the other hand, determination of the actual diameter of the wire is a difficult thing to do, and is so difficult that he would not plan to do it habitually in the applications of this method, but instead would simply employ the nominal diameter as a basis for his corrections, <u>then</u>, in evaluating the precision of this method, it would, I feel, be appropriate to compute a component of random error arising from deviations of actual diameters from the nominal in randomly chosen pieces of wire, and to incorporate this in his computed or estimated probable error of the procedure. This is a ticklish matter, and I do not want to go into it any further here, but I feel that it does deserve careful attention. No doubt each of you can think of instances of this sort in your own work.

THE NATIONAL BUREAU OF STANDARDS

Functions and Activities

The National Bureau of Standards is the principal agency of the Federal Government for fundamental and applied research in physics, mathematics, chemistry, and engineering. Its activities range from the determination of physical constants and properties of materials, the development and maintenance of the national standards of measurement in the physical sciences, and the development of methods and instruments of measurement, to the development of special devices for the military and civilian agencies of the Government. The work includes basic and applied research, development, engineering, instrumentation, testing, evaluation, calibration services, and various scientific and technical advisory services. A major portion of the NBS work is performed for other government agencies, particularly the Department of Defense and the Atomic Energy Commission. The functions of the National Bureau of Standards are set forth in the Act of Congress, March 3, 1901, as amended by Congress in Public Law 619, 1950. The scope of activities is suggested in the listing of divisions and sections on the inside of the front cover.

Reports and Publications

The results of the Bureau's work take the form of either actual equipment and devices or published papers and reports. Reports are issued to the sponsoring agency of a particular project or program. Published papers appear either in the Bureau's own series of publications or in the journals of professional and scientific societies. The Bureau itself publishes three monthly periodicals, available from the Government Printing Office: the Journal of Research, which presents complete papers reporting technical investigations; the Technical News Bulletin, which presents summary and preliminary reports on work in progress; and Basic Radio Propagation Predictions, which provides data for determining the best frequencies to use for radio communications throughout the world. There are also five series of nonperiodical publications: the Applied Mathematics Series, Circulars, Handbooks, Building Materials and Structures Reports, and Miscellaneous Publications.

Information on the Bureau's publications can be found in NBS Circular 460, Publications of the National Bureau of Standards ($1.00). Information on calibration services and fees can be found in NBS Circular 483, Testing by the National Bureau of Standards (25 cents). Both are available from the Government Printing Office. Inquiries regarding the Bureau's reports and publications should be addressed to the Office of Scientific Publications, National Bureau of Standards, Washington 25, D. C.

NBS