QUANTI-FICATION

A HISTORY OF THE MEANING OF MEASUREMENT IN THE NATURAL AND SOCIAL SCIENCES

Edited by Harry Woolf

QUANTIFICATION

A HISTORY OF THE MEANING OF

MEASUREMENT IN THE NATURAL

AND SOCIAL SCIENCES

Edited by Harry Woolf

THE **BOBBS-MERRILL** COMPANY, INC.
A SUBSIDIARY OF **HOWARD W. SAMS & CO., INC.**
Publishers • INDIANAPOLIS • NEW YORK

Printed in the United States of America

CONTENTS

MAIN ARTICLES

QUANTIFICATION

A HISTORY OF THE MEANING OF MEASUREMENT IN THE NATURAL AND SOCIAL SCIENCES

The Conference on the History of Quantification in the Sciences

Measurement has long been considered a hallmark of science properly practiced, and once a new discipline has developed a mathematical discourse, it has almost immediately laid claim, at least in the language of its most enthusiastic disciples, to the significant status—science! In the larger task of understanding the world, quantification or measurement may not only give precision to the perpetual dialogue between nature and the scholar, but it may also enable such a conversation to include an ever-growing portion of the manifold and the complex through the employment of increasingly higher levels of abstraction. Thus, in an age concerned not only with the fine analysis of nature, but also with the characteristics of its own methods and tools, the meaning of measurement in all science becomes of fundamental concern to theory and practice alike.

Towards this end then, an understanding of the measurer and the measured, the papers in this volume are directed. They are the result of an extraordinary conference held in the offices of the Social Science Research Council on 20-21 November, 1959, a conference whose origin goes back to 1955 and the founding of the Joint Social Science Research Council-National Research Council Committee on the History of Science. Taking precedent from the panel on the history of science established under Dr. Harry Alpert at the National Science Foundation, and with the cooperation of Dr. Detlev Bronk, then president of the National Academy of Science, the Joint Committee was created and first met in December 1955. The members of the original committee were I. B. Cohen, G. W. Corner (who later resigned to be replaced by R. B. Lindsay), H. Guerlac, M. H. Ingraham, R. K. Merton, H. L. Shapiro, G. R. Willey and R. H. Shryock as chairman.

In its meetings since 1955, the Joint Committee has been concerned with various aspects of the history and sociology of science, including the status of the field itself. In the spring of 1956 ideas were developed within the Committee for two conferences, which have since taken place: the first was devoted to the problems of interpretation in the history of science, and the second, at Dr. Shryock's suggestion, dealt with a composite analysis of the introduction and development of quantitative techniques in the natural and social sciences—the theme of the papers in this volume. The first conference took place at the University of Wisconsin on 1-10 September 1957. Under Professor Marshall Clagett's superb management, an excellent program developed, ranging in interest from the formation of the idea of the conservation of energy to problems in teaching the history of science. The papers and formal comments of this Wisconsin conference were assembled afterwards as *Critical Problems in the History of Science,* edited by Marshall Clagett and published by the University of Wisconsin Press, 1959.

Between 1956 and 1958, plans for the second conference to emerge from the deliberations of the Joint Committee matured, and Dr. Pendleton Herring, of the Social Science Research Council, requested and received the necessary support for this special meeting from the National Science Foundation. When it took place, the Conference on the History of Quantification in the Sciences was attended by some thirty scholars representing eleven academic disciplines: the history of science, physics, chemistry, biology, botany, mathematics, psychology, sociology, economics, political science and anthropology. The geographical and institutional distribution of the group was as varied as the professional interests of its members, for in addition to Professors Cohen, Guerlac, Lindsay, Merton and Shryock of the Joint Committee, the following took part: Harry Alpert, University of Oregon; Bernard Barber, Barnard College; Edwin G. Boring, Harvard University; Marshall Clagett, University of Wisconsin; A. C. Crombie, Oxford University (in residence at Princeton University at the time); Philip Frank, Harvard University; R. W. Gerard, University of Michigan; David R. Goddard, University of Pennsylvania; Mark Graubard, University of Minnesota; Earl J. Hamilton, University of Chicago; Pendleton Herring, Social Science Research Council; Alexandre Koyré, Institute for Advanced Study, Princeton; Simon Kuznets, The Johns Hopkins University; Thomas S. Kuhn, University of California, Berkeley; Paul F. Lazarsfeld, Columbia University; Daniel Lerner, Massachusetts Institute of Technology; Solomon Pines, Hebrew University, Jerusalem; Derek J. Price, Yale University; Albert C. Spaulding, National Science Foundation; Joseph J. Spengler, Duke University; M. H. Trytten, National Research Council; Charles F. Voegelin, Indiana University; S. S. Wilks, Princeton University; Harry Woolf, University of Washington.

The papers for each session were distributed to the participants in advance of the conference, and each of the four sessions was limited to a discussion of two papers. Instead of a chorus of separate and independent voices, the members of the conference discovered that not only had they much to say to one another, but upon occasion they could sing in close harmony indeed. Collaboration came naturally as the participants dealt with the problem of emerging quantification in different disciplines at different times, and almost from the very beginning the conference became a true symposium in the classic mold. The bracketed texts which appear in some of the articles and the appendices attached to others represent changes and additions made by the authors in response to the discussion of their papers at the conference.

Thus, these papers are presented to the reader with the hope that the intellectual gain registered by all who took part in the Conference on the History of Quantification in the Sciences may go beyond the offices of the Social Science Research Council where it took place. That it is possible is due to the generous cooperation of the authors involved, the patience and wise counsel of Dr. Herring and Dr. Shryock and the funds of the National Science Foundation. In the face of extraordinary conditions, the able assistance of Mrs. Carol B. Hewitt and Miss Dorothy Stratton helped to make the publication of these papers possible. For so much freely given, the editor of *Isis* can only express his warmest thanks.

Madras, India
18 April 1961

Harry Woolf, Editor of *Isis*

Some Aspects of Quantification in Science

*By S. S. Wilks**

IT is a difficult assignment for a person who is neither a philosopher nor a historian of science to fill the place of Professor Ernest Nagel in initiating discussion at this conference on quantification in the sciences. Professor Nagel, if he could have been here, would have performed these duties as a philosopher with broad experience and knowledge about the nature and history of quantification in the sciences. I can only approach the task as a mathematical statistician who has given some thought to the nature of quantification in science without much knowledge, however, of the history of quantification in the various sciences.

The subject of quantification in science is an enormous one with many aspects. The foundation of quantification is measurement, and any discussion of the nature of quantification must necessarily begin with a discussion of the nature of measurement. In this paper I shall not try to do more than to direct your attention to some of the basic concepts and requirements involved in measurement and quantification as we see them today, without attempting to trace the origin and development of these concepts historically. Having had an opportunity to read all of the manuscripts prepared for the Conference, I have observed that most of these concepts occur at least implicitly in concrete settings at various points in the manuscripts, and are thereby placed somewhat in historical perspective by the various speakers at this Conference.

The first requirement about measurement which should be mentioned is that making a measurement must be an *operationally definable process*. That is, a measurement process must be defined by specifying a set of realizable experimental conditions and a sequence of operations to be made under these conditions which will yield the measurement. The basic reason for such a requirement is to make the measurement process as objective as possible so that different competent scientists operating the process can obtain comparable results.

Making a measurement is an extremely widely applicable concept. At the lower end of the scale of complexity, counting and recording the number of rows of grains in an ear of corn or even counting and recording the number of heads one obtains if a single coin is tossed once are acts of making measurements. The existence of an ear of corn or a coin and an observer with an ability to count up to twenty and record his results are sufficient to make such a measurement.

At the upper end of the scale are highly complicated measurement processes

* Princeton University.

like determining the velocity of light or picking weak signals out of radio astronomy recordings loaded with background noise. While there are many measurement operations in the physical sciences which are complex from the point of view of instrumentation, and specification of measurement conditions, there are measurement operations in other fields of inquiry, particularly the social sciences, which are at least as complex in other ways. A good example is the procedure used by the Bureau of the Census[1] in estimating the number of employed and unemployed workers in the United States during a given week. In making such a measurement, great care is required in the selection, by certain randomization procedures, of a national sample of households to be contacted, in standardizing and controlling the interviewing procedures to be used by a nation-wide field staff, and in coding and analyzing the results. Stephan and McCarthy[2] have given a full discussion of measurement processes of this type.

It should be pointed out that operationally definable processes do not necessarily have to terminate in measurement-taking. For example, the manufacture of an article under mass-production conditions is an operationally definable process. Repeating such a process, however, would normally yield articles comparable with respect to the outcomes of any measurement process which might be made on them.

The second basic requirement of a measurement process is that of *reproducibility* of the outcome. Once a measurement process has been defined as objectively as possible, repeating the process should yield measurements in "reasonable agreement" with each other. That is, two competent scientists performing the measurement process independently, should obtain determinations in "reasonable agreement" with each other. The more objectively and accurately the measurement process is defined, the closer, in general, is the agreement to be expected. This property of reproducibility of a measurement process is sometimes called the *reliability* or the *precision* of the process. There are many conditions, of course, in which it is impossible to repeat a measurement process in a strict sense. For example, if conditions change markedly in time, repeatability may involve repetitions of the measurement process as nearly simultaneously as possible. Or if the object being measured is destroyed in the process of making the measurement, as often occurs in making life tests and other critical tests on mass-produced articles, repetition would involve applying the measurement process to a second article from the mass-produced lot.

The requirement of reproducibility of a measurement process is not always properly appreciated even in the older branches of science, not to mention the newer ones. There is often great temptation to *assume* that the measurement process is so carefully and objectively defined and controlled that determinations which would result from two independent runs of the measurement process will inevitably be in "reasonable agreement," and hence that there is

[1] United States Bureau of the Census (1954), *The Measurement of Employment and Unemployment by the Bureau of the Census in its Current Population Survey*, Report of Special Advisory Committee on Employment Statistics.

[2] Frederick F. Stephan and Philip J. McCarthy, *Sampling Opinions* (New York: John Wiley and Sons, 1958).

no need of repeating the measurement process. This is one of the most hazardous assumptions which can be made in any field of science. An excellent discussion of many of the factors which can upset such an assumption in scientific research has been given by Wilson.[3]

When results of scientific research have practical application and when the measurement procedures pass over into the hands of practitioners, the concept of reproducibility is frequently completely abandoned, unless it is incorporated into procedural doctrine. For example, it has been found by Chiang, Hodges and Yerushalmy[4] that diagnoses of tuberculosis based on chest x-ray negatives have low reliability in medical practice. The unreliability apparently stems from the subjectivity with which the clinician interprets what he sees on the negative. If the chest x-ray negatives carry sufficient information for correct diagnosis of tuberculosis, the unreliability is due to failure to define sufficiently objectively what the clinician must do to elicit this information. If, as is more likely the case, they carry only partial information for correct diagnosis, then there is a definite limit to the reliability which can be achieved.

The third requirement of basic importance for a measurement process is that of the *validity* or the *accuracy* of the process: that is, the extent to which the process yields "true" measurements of the object being measured. The notion of reproducibility or reliability or precision is not to be confused with that of validity or accuracy of a measurement process. A satisfactory measurement process requires high reliability and high validity. There can be situations in which the measurement process has high reliability but low validity. For example, if the sight on a good rifle is not properly aligned, a good rifleman with a steady hand would be expected to achieve high reliability but low validity with his marksmanship. That is, he could place successive shots close together but *not* around the bull's eye. To achieve high reliability and also high validity he would have to place the shots close together *and* around the bull's eye, a result to be expected from an expert rifleman and a good rifle with its sights properly aligned.

In the example of chest x-ray diagnoses for tuberculosis, high consistency between the conclusions of two clinicians working independently on a large number of x-ray negatives would not imply validity of their diagnoses. But high validity would imply that both clinicians working independently would agree on their diagnosis from nearly all films, and furthermore nearly all films diagnosed as positive would be from persons infected with tuberculosis while nearly all films diagnosed as negative would be from persons not infected.

The problem of achieving validity by a measurement process is usually much more difficult than that of achieving reliability. For reliability merely requires reproducibility of two or more repetitions of the measurement process, while validity requires that the numerical value produced by the measurement process be approximately the same as the true value of the quantity being measured as determined by some independent and valid procedure. In many

[3] E. Bright Wilson, Jr., *An Introduction to Scientific Research* (New York: McGraw-Hill Book Company, 1952).

[4] C. L. Chiang, J. L. Hodges, Jr. and J. Yerushalmy, "Statistical Problems in Medical Diagnosis," Vol. IV, *Third Berkeley Symposium* (Berkeley: University of California Press, 1956).

situations such an independent procedure does not exist, and hence the problem of establishing validity in the true sense of the word cannot be solved. In such situations the case for validity is made to rest on the logic and experimental rigor of the measurement process together with cross checks of the measurement results with independently acquired facts. For example, Michelson's[5] procedure for measuring the velocity of light yielded quite highly reproducible results from trial to trial, but the validity of his estimate of the velocity of light is allowed to rest on the fact that the use of his estimate in describing physical phenomena involving the velocity of light does not lead to significant contradictions or inconsistencies. To take an example from the social sciences, we may ask whether the current Bureau of the Census procedure for measuring (estimating) the number of employed and unemployed persons in the United States from 35,000 households each month is valid, that is whether the values obtained are "reasonably close" to the true values of the number of unemployed in these months. The problem of validity here is extremely difficult. The case for validity is made partly on the logic of the design of the sampling system and the control exercised in the execution of the design and analysis of the results, and partly on the precision with which various statistical quantities known from previous censuses and surveys can be estimated from the sample results.

It should be pointed out that there are situations in which the purpose of a measurement process is to provide a sort of index for which the concept of validity has meaning only in some general and usually unmeasurable sense. But in such a situation the importance of the requirement of reliability is in no way diminished. An examination in a given subject taken by a group of individuals is an example of such a process. The purpose of the examination is to provide scores or indices on the individuals indicative in some general sense of how much they know about the subject. The notion of a "true" score for an individual is useful for conceptual purposes, but it is unmeasurable. Even so, the examination can and should possess as much reliability as possible, that is, if a second similar examination in the subject is given to the group of individuals the ranking of their scores on the second examination should be in "reasonable agreement" with the ranking of their scores on the first examination. For a thorough discussion of reliability of an examination, the reader is referred to Gulliksen[6] who also discusses the problem of evaluating the validity of an examination. Procedures for determining values of cost of living indices, economic indicators, and measures of the effectiveness of competing weapon systems are further examples of such a measurement process.

There is a class of highly practical measurement processes in which validity is crucial and verifiable, which we may call calibration processes. In a calibration process a scale of values of an auxiliary but easy-to-measure variable y is constructed so that the values of y corresponding to various specified values of the basic but difficult-to-measure variable x are determined experimentally.

[5] A. A. Michelson, E. G. Pease, and F. Pearson, "Measurement of the velocity of light in a partial vacuum," *Astrophys. J.*, 1935, *82:* 26-61.

[6] Harold Gulliksen, *Theory of Mental Tests* (New York: John Wiley and Sons, 1950).

These pairs of corresponding values of x and y are then used to construct by interpolation a scale of y values corresponding to all x values in the range of interest. Then to each value of y selected by an indicator would correspond a value of x. Most instrumentation dials are applications of the calibration process. An automobile speedometer is a typical example. Values of y in miles per hour are printed at appropriate points along the y scale. A pointer activated by the speed of the automobile would indicate a value of y in miles per hour. Validity of the speedometer thus requires that the speed of the automobile as indicated by the pointer on the y scale must agree "within practical limits" with the true speed x of the automobile as determined by any valid method independent of the automobile's own speedometer.

Let us now turn to some quantification concepts concerning aggregates or systems of measurements. It must be emphasized first of all that the quality of an aggregate or system of measurements depends on the quality of individual measurements which, in turn, depend on the degree to which the underlying measurement process satisfies the three fundamental requirements already discussed, namely that the process be operationally definable, that it yield reliable measurements, and that it yield valid measurements.

Perhaps the simplest kind of an aggregate of measurements is one generated by applying a measurement process to each member of a *sample* of objects drawn from a *population* of such objects. The main purpose of such an aggregate of measurements is to learn something about the variation of the measurements from object to object in the sample and to estimate the mean, or some other function, of the measurements one would obtain if all objects in the population were subjected to the given measurement process. Measurements on samples to make estimates of population characteristics are widely used in science and technology. For example, samples of articles from mass-produced lots are widely used for estimating quality characteristics of the entire lots. Samples of households are used for estimating characteristics of the population of households in a city, county, state or the entire nation.

In order for a sample to provide a scientific basis for estimating the mean or other parameters of the population of interest to the investigator, special attention must be given to the process of drawing the sample from the population. Procedures for drawing such samples are based on *randomization principles,* which have been highly developed and widely used in many branches of science. The reader will find a thorough discussion of these principles in Cox.[7]

Principles of sampling are also used in comparative experiments. In the simplest kind of a comparative experiment, two random samples are drawn from a given population of objects. The objects in one sample are held as controls and the objects in the other are subjected to the treatment under study. The given measurement process is then applied to all objects in both samples. The mean or some other quantity is computed from each sample, and a significance test based on probability theory is used to determine whether the measurements in the two samples are behaving like measurements in two random samples from identical populations. If the hypothesis of identical

[7] D. R. Cox, *The Planning of Experiments* (New York: John Wiley and Sons, 1958).

populations is strongly contradicted by the significance test as applied to the measurements, it is concluded that the treatment has a definite effect which can then be estimated from the measurements.

Such a comparative experimental procedure is used widely in many branches of science and technology, but hardly at all in others. One still finds hazardous conclusions in scientific papers typified as follows: that a certain treatment produced an effect based on an experimental set-up involving a small control group and a small experimental group, when as a matter of fact, differences between the means of the two groups at least equal in magnitude to the observed difference could have occurred with relatively high probability under the assumption that the treatment had no effect whatever.

A comment should be made here about the importance of using the principle of randomization in drawing the two samples involved in a comparative experiment. If the two samples are selected from the population by any other principle than randomization, significance testing based on probability theory becomes inapplicable, and the possibility of other factors than the treatment under study cannot be ruled out as the cause of observed differences between the two samples. The current controversy over whether heavy cigarette smoking causes lung cancer hinges on this point. In the usual statistical studies of this problem, incidence of lung cancer in a large sample of heavy smokers is compared with that in a large sample of non-smokers, and is found to be larger. But since the two samples of persons are self-selected and not selected at the outset by principles of randomization, the possibility cannot be eliminated that persons who elect to smoke heavily may be constitutionally more susceptible to cancer (and other diseases) than those who do not smoke. As a matter of fact, it has been pointed out by Berkson[8] and others that heavy cigarette smokers have not only a greater incidence of lung cancer than non-smokers, but also a greater incidence of circulatory and other kinds of disorders.

The quantification of inference procedures from random samples to populations has been developed for a very wide range of experimental situations. Experimental designs have been developed for studying effects of several factors or treatments simultaneously. Highly specialized analysis of variance procedures have been devised for decomposing total variability of measurements in a sample into components and identifying causes of various components of this variability, thus making it possible to reduce variability by eliminating some of the causes. Modern industrial quality control methods originated by Shewhart[9] and further developed by others are founded largely on procedures of this type.

I would now like to turn to other systems of measurements. It will be recalled that the first basic requirement for a measurement process involves the specification of a set of conditions under which the measurement process is to be operated. Very often the specification of a set of conditions amounts to holding important *independent variables* fixed at certain values and then op-

[8] Joseph Berkson, "Smoking and Lung Cancer: Some Observations on Two Recent Reports," *J. Amer. Statist. Ass.*, 1958, *53:* 28-37.

[9] Walter A. Shewhart, *Economic Control of Quality of Manufactured Product* (New York: Van Nostrand, 1931).

erating the measurement process. If a measurement is made one or more times for each of a number of combinations of values of these independent variables, then, by suitable interpolatory smoothing, one may find the quantity being measured as a function of these variables. Boyle's law relating pressure and volume of a gas at a fixed temperature is an example of such a system of measurements. Thus, at a fixed temperature if the volume v of gas is measured for each of a set of fixed values of pressure p, we obtain a set of points in the pv plane which is closely fitted by a curve having an equation of form $v = c/p$ where c is a constant depending on temperature, thus giving v as a function of p. If we have several independent variables, which for simplicity we keep to two, namely x_1, x_2, and obtain measurements on a quantity q for various combinations of values of the variables x_1, x_2, we obtain points in the 3-dimensional space of q, x_1, x_2. If a smooth function of form $f(x_1, x_2)$ is fitted to these points by some statistical technique such as least squares, we obtain a smooth ***regression surface*** for q on x_1, x_2 having an equation $q = f(x_1, x_2)$ which "fits" the *observed* points, that is the points obtained by measurement. If the observed points fall "reasonably close" to this fitted regression surface, we then have a model useful for estimating the value of q for any point (x_1, x_2) in the domain covered by the experiment. If q_1, q_2, $\ldots$, q_n are the observed values of q corresponding respectively to n combinations of values of (x_1, x_2), say (x_{11}, x_{21}), (x_{12}, x_{22}), $\ldots$, (x_{1n}, x_{2n}), then $q_1 - f(x_{11}, x_{21})$, $q_2 - f(x_{12}, x_{22})$, $\ldots$, $q_n - f(x_{1n}, x_{2n})$ are the "errors" of the observed points with respect to the regression surface. The larger these errors in magnitude, the lower the quality of this regression surface for estimating q for given values of (x_1, x_2).

Regression functions of the type discussed above which may involve several independent variables $x_1, x_2, \ldots, x_k$ are used for estimating the value of q for given values of $x_1, x_2, \ldots, x_k$ in many branches of science and technology. In the particular case where the regression function is satisfactorily approximated by a linear form such as $\beta_0 + \beta_1 x_1 + \ldots, + \beta_k x_k$, the problem of fitting it to the observed points involves the solution of a system of linear equations in the unknown β's. The standard procedure for determining the β's is by the method of least squares which yields a matrix for the set of equations whose elements are readily computable from the observations. As a matter of fact, the solution of sets of linear equations of this kind in many variables is now a routine matter for high speed digital computors.

It should be noted from the preceding discussion that regression functions are essentially mathematical models for estimating values of the quantity q for specified values of the independent variables, $x_1, x_2, \ldots, x_k$ for which no observed value of q was obtained. While this type of mathematical model, that is the function $f(x_1, x_2, \ldots, x_k)$ chosen to be fitted to the observed points, is rather empirical, it is often remarkably realistic in terms of making useful estimates or predictions of measurement outcomes for all combinations of values of the independent variables in the domain of study.

Finally, I would like to comment on what we may regard as perhaps the highest form of quantification in science. This consists of the mathematical models which describe the essential features of the quantitative relationships inherent in vast systems of measurements. A model of this type not only provides a relatively simple and elegant scheme for describing a system of measurements which have already been made but also serves as a dependable instrument for predicting the outcome of further measurements which would belong to the system if they were made. Examples of such models are Kepler's laws of planetary motion, Newton's more general laws of motion, the focal laws of optics, differential equations of fluid mechanics, Mendelian laws of genetics, laws of probability for independent events, the Weber-Fechner law in psychophysics, and so on.

A great deal of effort is expended in devising and testing mathematical models based on specified assumptions in an attempt to describe systems of measurements. Frequently, models are devised which describe "reasonably well" existing sets of measurements but which do not stand the test of validity in satisfactorily predicting the outcome of further measurements. Many branches of science and technology abound in mathematical model-building activity done purely on the basis of assumptions, and often with little, if any, experimental knowledge to go on. Such model-building usually occurs as a "theoretical analysis" done in connection with the study of alternative designs of devices and systems as a guide to the selection of a design from the alternatives in advance of the construction of the device or system. For example, development of the "Mousetrap" device early in World War II for throwing barrages of small contact depth charges forward in attacking a submarine was preceded by a great deal of probability analysis to compare its expected effectiveness with that of conventional "ash can" depth charge attack procedures. Mathematical model-building as an effort to provide a description of a system of measurements or observations to be expected if they were made, but which is never actually followed up and tested against experimental results, must be regarded as a mathematical exercise rather than as scientific quantification.

Quantification in Medieval Physics

*By A. C. Crombie**

A WORTH-WHILE discussion of quantification in medieval physics requires particular care in deciding what is to be talked about. The whole question is obviously much less clear and much more equivocal in this period than it became later. So it is important to begin with some distinctions. I shall distinguish first between quantified procedures and quantified concepts, and I shall take a quantified procedure in science to be one that aims at measurement, that is, any procedure that assigns numbers in a scale. To be complete such a procedure must comprise both mathematical techniques for operating the scale theoretically and measuring techniques for using it to explore the world. Technology need contain little more than procedures of these kinds, which provide for the measurements and calculations with which it is concerned. But most sciences aim beyond these at providing explanations by means of a system of theory. So a quantified science, as distinct from quantified technology, comprises not only quantified procedures but also quantified explanatory concepts, each applicable to the other within a theoretical system. The development of a science then takes place through a dialogue between its theories and its procedures, the former offering an exploration of the expected world through predictions and explanations made by means of the technical procedures, and the latter confronting these theoretical expectations with the test of quantified data.

A dialogue of this kind requires that both sides should speak the same language. We are so familiar with the close and precise adaptation of conceptual and procedural language to each other in modern physics that it may come as a surprise to find authentic scientific systems in which this is not the case. Yet we do not have to look very far to find examples. In the contemporary social sciences and in psychology, they are notorious. We do not have to go many decades back in the history of modern genetics to find a very incomplete and interrupted dialogue between theories and procedures. Somewhat earlier, in the eighteenth century, we find the same situation in chemistry. The main interest of medieval physics in this context seems to me to be that it provides the earliest example in the development of modern science in which we can study the state of affairs when the dialogue between concepts and procedures was incomplete or absent. Then we can study the difference it made when clear and exact communication was opened, as it was in the seventeenth century. I shall assume that it is my brief to discuss medieval physics as a case history of a general problem. At the same time, I shall as-

* Princeton University; Oxford University.