



MEASUREMENT, THEORY OF

by (in part)

R. Duncan Luce

**from
the
fifteenth
edition
of
ENCYCLOPAEDIA
BRITANNICA**

©1974

by Encyclopædia Britannica, Inc.

Copyright under International
Copyright Union

All rights reserved under Pan American
and Universal Copyright Conventions

by
Encyclopædia Britannica, Inc.
Printed in U.S.A.

HELEN HEMINGWAY BENTON, Publisher

the deviation IQ, mainly because the spread around the average has not been uniform due to different ranges of item difficulty at different age levels. The deviation IQ, a type of standard score, has a mean of 100 and a standard deviation of 16 for each age level. Current practice with the Stanford-Binet test reflects the finding that average performance on the test does not increase beyond age 18. Therefore, the chronological age of any individual older than 18 is taken as 18 for the purpose of determining his IQ.

ASSESSING TEST STRUCTURE

Factor analysis. Test items that measure the same characteristic tend to be positively correlated with one another. A British psychologist, Charles Edward Spearman, systematically explored such intercorrelations to provide evidence that much of the variability in scores that children earn on tests of intelligence depends on a general underlying factor that he called *g*. In the United States, Thurstone developed a statistical technique called multiple-factor analysis, with which he was able to demonstrate additional (special rather than general) factors that tend to crystallize as people grow older. He provided evidence that, among adults, one specific talent (such as verbal comprehension) is distinguishable from another (*e.g.*, numerical ability).

Early computational methods in factor analysis have been supplanted by mathematically more elegant, computer-generated solutions; and the general technique is widely applicable beyond the field of psychological measurement.

Rooted in extensive applications of factor analysis, a structure-of-intellect model developed by an American psychologist, J.P. Guilford, posited a very large number of factors of intelligence. Guilford envisaged three intersecting dimensions corresponding to four kinds of test content, five kinds of intellectual operation, and six kinds of product. Each of the 120 cells in the cube thus generated was hypothesized to represent a separate ability, each as a distinct factor of intellect. Educational and vocational counselors usually prefer a substantially smaller number of scores than the 120 that are implied by this model.

Alternatives to factor analysis. Especially with respect to personality measures, simple techniques called cluster analysis have been proposed as alternatives to the statistical complications of factor analysis. In one form of cluster analysis, tables of correlation coefficients are simply inspected in an effort to find sets (clusters) of items or test scores that show high relationships among themselves.

Early forms of cluster analysis introduced in the 1930s required laborious visual inspection and time-consuming computations. With the advent of high-speed computers, new methods proliferated rapidly. Although it has been argued that clusters bear a closer relation to reality than do statistically defined factors, the correspondence appears to be at a descriptive rather than at an interpretive level. Taken separately or together, various types of cluster analysis fail to yield equivalent results; a number of authorities thus regard them as adjuncts to, rather than substitutes for, factor analysis.

Techniques that are quite similar to the usual factor-analytic approach but provide different insights are those related to inverse factor analysis (based on intercorrelations among persons instead of among tests). One such method begins with what is called a *Q-sort*, in which a subject sorts words or other stimuli into categories on the basis of his preference so that they form a bell-shaped (normal) distribution.

This method is widely applicable to items from personality tests; first, correlations of data obtained from two persons' *Q*-sorts of the same items are computed. The inverse factor analysis of a matrix of such intercorrelations yields factors indicative of common ways people have of reacting to stimuli.

Another popular technique (the semantic differential) for detecting underlying variables was offered by Charles E. Osgood, an American psychologist. In this method,

the name of any object or concept (*e.g.*, religion or piety) is presented to an individual, who rates it on a series of seven-point scales defined semantically by pairs of adjectives (such as "good" versus "bad" or "valuable" versus "worthless"). When different concepts receive similar ratings on the semantic scales, they show evidence of representing the same underlying factor. The method has been applied mainly to psychological problems, involving attitudes.

Profile analysis. With the fractionation of tests (*e.g.*, to yield scores on separate factors or clusters), new concern has arisen for interpreting differences among scores on the underlying variables, however conceived. Scores of an individual on several such measures can be plotted graphically as a profile; for direct comparability, all raw scores may be expressed in terms of standard scores that have equal means and variabilities. The difference between any pair of scores that have less than perfect reliability tends to be less reliable than either, and fluctuations in the graph should be interpreted cautiously. Nevertheless, various features of an individual's profile may be examined: scatter (fluctuation from one measure to another) and relative level of performance on different measures. (The particular shape of the graph, it should be noted, partly depends upon the arbitrary order in which measures are listed.) One may also statistically express the degree of similarity between any two profiles. Such statistical measures of pattern similarity permit quantitative comparison of profiles for different persons, of profiles of the same individual's performance at different times, of individual with group profiles, or of one group profile with another. Comparison of an individual's profile with similar graphs representing the means for varied occupational groups, for example, is useful for vocational guidance or personnel selection.

Prospects. Despite progress over the last century, psychological measurement warrants improvement. Brief appraisals of individual potentiality—whether in educational, vocational, or personal adjustment settings—are recognizably far from perfect. The complexities of human behaviour attributable to the myriad effects of unanticipated combinations of genes, prenatal influences, and vagaries of environmental effects make it surprising that a relatively brief test or even a full day of testing can predict future behaviour to any extent at all beyond a chance level.

Predictions are successful beyond chance expectations, however. And clearly the forecasts of psychometricians and other specialists in educational and psychological measurement have much to commend them in contrast with subjective appraisals of sooth-sayers, phrenologists, astrologers, or untrained personnel interviewers who happen to like people.

BIBLIOGRAPHY. DOROTHY ADKINS WOOD, *Test Construction* (1961), a simplified treatment of measurement principles, rules for test construction, and statistical techniques; ANNE ANASTASI, *Psychological Testing*, 3rd ed. (1968), an authoritative text and reference book with emphasis on current psychological tests; L.J. CRONBACH, *Essentials of Psychological Testing*, 3rd ed. (1970), another modern and insightful text and general reference; J.P. GUILFORD, *Psychometric Methods*, 2nd ed. (1954), a widely used book that attempts to integrate psychophysical scaling and psychological measurement methods; HAROLD GULLIKSEN, *Theory of Mental Tests* (1950), a basic theoretical reference; H.H. HARMAN, *Modern Factor Analysis*, 2nd ed. rev. (1967), an eclectic treatment of factor-analytic theory and methods; PAUL HORST, *Psychological Measurement and Prediction* (1966), a discussion of practical requirements of psychological measurement as well as of technical problems in prediction; F.M. LORD and M.R. NOVICK, *Statistical Theories of Mental Test Scores* (1968), a highly technical presentation; R.L. THORNDIKE (ed.), *Educational Measurement*, 2nd ed. (1970), contains specially prepared chapters by authorities in particular fields of measurement.

(D.C.A.)

Measurement, Theory of

Measurement theory studies the practice of associating numbers with objects and empirical phenomena. It attempts to understand which qualitative relationships lead to numerical assignments that reflect the structure

of these relationships, to account for the ways in which different measures relate to one another, and to study the problems of error in the measurement process.

HISTORY AND NATURE OF MEASUREMENT THEORY

History of magnitudes theory

An early and relatively abstract theory of magnitudes was developed in the 4th century BC by two mathematicians, Eudoxus of Cnidus, probably a member of Plato's Academy, and Theaetetus, author of the book on the regular solids and one of those on irrational numbers in Euclid's *Elements*, in which their theory of magnitudes was given its fully developed presentation. Theirs was a highly sophisticated theory of measurement. Historical evidence suggests that it was developed for purely theoretical purposes, to account for incommensurable magnitudes—pairs of numbers, such as 1 and the square root of 2, for which no third number exists of which both are even (or integral) multiples—rather than to solve any practical problems, such as those of surveying. Problems of measurement, of both a theoretical and a practical sort, continued to be discussed throughout ancient times and during the Middle Ages.

In modern times, Euclid's view of the theory of magnitudes was still relevant as late as the 18th century; it appeared, for example, in Isaac Newton's *Arithmetica Universalis* (2nd ed., 1728). The modern, axiomatic treatment of the theory—i.e., its logical development from fundamental concepts and axioms—can be dated from a famous paper by Herman Ludwig Ferdinand von Helmholtz, a German philosopher and wide-ranging scientist, entitled "Zählen und Messen erkenntnistheoretisch betrachtet" (1887; Eng. trans., "Counting and Measuring," 1930). Considerable impetus was given by a classical work of Ludwig Otto Hölder, a German mathematician, entitled "Die Axiome der Quantität und die Lehre vom Mass" (1901; "The Axioms of Quantity and the Theory of Measure"), in which he provided two sets of axioms, one for the measurement (or, as he said, "the quantity") of additive attributes, such as mass, which can be built up by addition of parts to form a total bulk, and another set for nonadditive attributes involving the measurement of intervals, as in temperature or longitude. The tradition of Helmholtz and Hölder has been of importance primarily in the physical sciences.

The consideration of measurement in the social sciences, especially in economics, also has had a long history. As early as the 18th century, Jeremy Bentham, a British Utilitarian moralist, in a treatise, *The Principles of Morals and Legislation* (1789), attempted to provide a theory for the measurement of utility, or value. In the 19th century there was an effort to establish that much of economic theory could be developed solely on the basis of ordinal preferences among the economic goods being studied, without any more demanding kind of numerical measurement. The work of Vilfredo Pareto, an Italian economist and sociologist, entitled *Manuale di economia politica* (1906; "Handbook of Political Economics"), was especially important in this connection. In the mid-20th century the study of measurement in economics and in psychology was stimulated by the theory of utility set forth in *Theory of Games and Economic Behavior* (1944) by the mathematician John von Neumann, a scientist of unusual breadth, and the economist, Oskar Morgenstern. It makes use of the formation of so-called risky alternatives (see below *Economics and psychology*).

The three basic problems of measurement theory

An examination of specific examples of axiomatic measurement can lead to a general view of measurement theory. The first problem for any such theory is to justify the assignment of numbers to objects or phenomena—to pass from empirical procedures and operations to a numerical representation of these procedures. From an axiomatic standpoint, the problem—known as the representation problem—is first to characterize the formal or abstract properties of these procedures and observations and then to show mathematically that these axioms permit the construction of a numerical assignment in which familiar abstract relations and operations, such as "is greater than or equal to" (symbolized \geq) and "plus"

(+), correspond structurally to the empirical (or concrete) relations and operations.

The second fundamental problem is that of the uniqueness of the representation—i.e., how close it is to being the only possible representation of its type. The representation of mass, for example, is unique in every respect except the choice of unit; e.g., the representation is different for pounds than for grams or grains. Ordinary measurements of temperature, however, are unique in everything except the choice of both unit and origin—the Celsius and Fahrenheit scales differ not only in the size of unit but also in the zero point. A U.S. psychophysicist, S. Smith Stevens, was among the first to place great emphasis on the uniqueness of measurement in relation to the problem of units and the consequences that this uniqueness has for data handling.

A subject closely related to the uniqueness of physical measurement is dimensional analysis, which, in a broad sense, attempts to explain why the various physical measures exhibit simple relations in the fundamental equations of classical physics. If length, time, and mass are taken to be the fundamental dimensions of mechanics, for example, then all other quantities, such as force or momentum, can be represented simply as products of powers of these dimensions—a fact that has strong implications for the forms of physical laws (see below *Universal, system, and material constants*).

A third central problem for the theory of measurement is error. In spite of the early development of precise astronomical measurement, no systematic theory of observational error appears to have been developed until the 18th century, first in an article by Thomas Simpson, an English mathematician, in 1757, and then in the fundamental work of two French mathematicians, Joseph-Louis, comte de Lagrange, who worked on the theory of numbers and celestial mechanics, and Pierre-Simon, marquis de Laplace, also an astronomer, famous for his fundamental work, *Traité de mécanique céleste* (1798–1827; *Celestial Mechanics*, 1966). Today, any really significant physical measurement is routinely reported together with some indication of the probable error, and theories are tested by confirming them within the errors of measurement. In the classical tests of Albert Einstein's general theory of relativity, for example, the discrepancies between predictions and observations fell mainly within the estimated errors of measurement (for further discussion of errors of measurement, see below *The problem of error*).

AXIOMATIC BASIS OF MEASUREMENT

Axiom systems for measurement differ in the amount and type of structure involved. All include an ordering relation, but that alone is not enough to formulate many scientific laws because its numerical representation is not closely prescribed. Additional structure and an increased uniqueness of the representation arise either by having an empirical operation of addition (extensive measurement), or by having entities that have several independent components (difference and conjoint measurement), or by other primitives that lead to a geometric representation.

Axioms of order. Whenever a measurement is made, it is done in such a way that the order induced on the objects by the assigned measure is the same as that obtained by the basic empirical operation in question. Such measures are said to be order preserving. Of course, every numerical inequality involved in measurement is transitive (i.e., such that $x \geq y$ and $y \geq z$ implies that $x \geq z$) and connected (i.e., either $x \geq y$ or $y \geq x$). It is therefore necessary that the empirical ordering—symbolized \geq (with \sim instead of $=$ to suggest real entities)—be transitive and connected in order for a representation of it to be possible by means of \geq . Relations exhibiting these two properties are called weak orders, in which "weak" simply means that indifference (symbolized \sim)—in which $x \sim y \equiv x \geq y$ and $y \geq x$ —is not necessarily equality (using \equiv to mean "if and only if"). Not every weak order, however, has an order-preserving numerical representation; the lexicographic

Axioms for binary relations

order in the plane $(x,y) \geq (x',y') \equiv x > x'$ or $x = x'$ and $y > y'$, is a counterexample. A second, independent property of numbers, which must in turn be reflected in the empirical ordering, is that rational numbers—*i.e.*, numbers expressible as the quotient of two integers—are order-dense (*i.e.*, they are such that between any two distinct real numbers lies a rational number) and countable (*i.e.*, they can be placed in one-to-one correspondence with the integers).

In 1895 it was shown by Georg Cantor, a German mathematician, that, for any empirical ordering, the existence of a numerical representation that preserves its order is equivalent to the ordering's being a weak one that includes in its domain a countable, order-dense subset. Any two such representations are so related that one can be mapped upon the other in a strictly increasing fashion.

Though various ordinal categorizations are widely used (*e.g.*, brightness of stars, magnitude of earthquakes, hardness of minerals), the fact that equivalent ordinal representations are seldom in simple proportion to one another makes them unsatisfactory in the statement of many scientific laws. To increase the uniqueness of the representation, some structure in addition to order must be preserved by the representation.

Axioms of extension. For many physical attributes—including mass, length, time duration, and probability—the objects or events exhibiting the attribute may be combined, or concatenated, to form new objects or events that also exhibit the attribute. Both alone and combined, for example, objects have mass. Denoting the concatenation of a and b by $a \circ b$, the assignment (designated ϕ) of a given set of numbers to the objects is called an extensive representation of the empirical ordering \geq and the concatenation \circ , provided that it is not only order preserving but also additive in the sense that $\phi(a \circ b) = \phi(a) + \phi(b)$.

Theories that are intended to account for the existence of such numerical representations must state empirical laws about ordering and concatenation separately, as well as how they interrelate. The key property of an empirical ordering is that it is a weak order, and that of a concatenation is that it is insensitive to the order of combination—concatenation is weakly commutative (*i.e.*, its order can be reversed), $a \circ b \sim b \circ a$, and weakly associative (*i.e.*, its terms can be regrouped), $(a \circ b) \circ c \sim a \circ (b \circ c)$ —in which \sim indicates an equivalent formulation. The key property relating them, weak monotonicity, is that the empirical ordering $a \geq b$ holds if and only if the ordering of its concatenation with any object c , $a \circ c \geq b \circ c$, also holds. As in ordinal measurement, a further and more subtle property, called Archimedean, is needed; this asserts that the elements within the structure are commensurable with one another. This property is formally analogous to the numerical Archimedean property that if $x > y > 0$, then for some integer n , $ny \geq x$ (meaning y is not “infinitesimally” small relative to x). With these stipulations, the numerical representation can be constructed.

The basic idea of the construction is simple and, in somewhat modified form, is widely used to carry out fundamental measurement. The measurer first chooses some object u , such as a foot rule, to be the unit of measurement. For any object a (say a house stud), he then finds other objects (studs) that are equivalent to a in the attribute in question (length). He concatenates these by laying them end-to-end in a straight line. Denoting by na the concatenation of any n of them, he now finds how many copies of u , say $m(n)$, are needed to approximate na ; *i.e.*, $[m(n) + 1]u \geq na \geq m(n)u$. The limit toward which $m(n)/n$ tends as n approaches infinity exists and is defined to be $\phi(a)$, which is the measurement sought; it can be shown to be order preserving and additive. In contrast to ordinal measurement, such an extensive representation is almost unique: it is determined except for a positive multiplicative constant or coefficient setting the scale; or, what is the same thing, only the choice of the unit is arbitrary. As a result, such a representation is called a ratio scale.

Sequences such as $na, n = 1, 2, \dots$, being ubiquitous in measurement practice, are called standard sequences. Practical examples are standard sets of weights in multiples of a gram and the metre rule, subdivided into millimetres.

For some purposes, including the development of other kinds of measurement—*e.g.*, difference and probability measurement—it is necessary to generalize the theory to cover concatenation operations (\circ) that are defined only for some pairs of objects and not for others. Probability, for example, is not additive over all pairs of events, but only over those that are disjoint. Such theories, developed during the 1950s and 1960s, all assume that certain empirical inequalities have solutions. One form of such solvability is: if $a > b$, then there exists a c such that c and b can be concatenated and $a \geq c \circ b$.

Axioms of difference. Length, but not mass, may be treated also in terms of intervals (a, b) on a line. Here the empirical ordering \geq is a quaternary rather than a binary relation, since it involves the comparison of two pairs, and so four points, such as the end points of two intervals; and concatenation is defined only for adjacent intervals; *i.e.*,

$$(a,b) \circ (b,c) = (a,c).$$

In addition to weak ordering, solvability, and Archimedean properties, various other axioms, depending on the exact representation to be constructed, must be satisfied. Two important representations are the algebraic difference one, in which an empirical ordering holds if and only if the differences of its elements form an ordered representation; *i.e.*,

$$(a,b) \geq (c,d) \equiv \phi(a) - \phi(b) \geq \phi(c) - \phi(d),$$

for which a key axiom is that the ordering is preserved through an interchange of arguments; *i.e.*,

$$(a,b) \geq (c,d) \equiv (a,c) \geq (b,d),$$

and the absolute difference representation, in which an empirical ordering holds if and only if the absolute values (symbolized $|\dots|$) of such differences are ordered; *i.e.*,

$$(a,b) \geq (c,d) \equiv |\phi(a) - \phi(b)| \geq |\phi(c) - \phi(d)|,$$

for which key axioms are that any interval is at least as large as the null interval; *i.e.*, $(a,b) \geq (a,a)$ and that an interval is unordered, or the same when measured “backwards”; *i.e.*, $(a,b) \sim (b,a)$.

Such representations are called interval scales, because they are unique in every respect except for the possibilities of multiplication by a positive constant and addition of a constant; *i.e.*, both the unit and zero of the measure are arbitrary, and the equality of intervals is preserved through all transformations generated by altering these two scale factors.

Axioms of conjointness. A major hindrance to the development of fundamental measurement, other than probability, outside physics has been the failure to uncover suitable empirical concatenation operations for attributes such as utility, loudness, intelligence, or hunger. An alternative approach, implicit in much derived measurement in classical physics, has recently been made completely explicit in the behavioral sciences. It rests on the fact that many—indeed, most—attributes are defined for entities having at least two independent components, each of which affect the attribute: kinetic energy and momentum, for example, are both affected by mass and velocity; mass is varied by both changing density and volume; preference among gambles is manipulated both by outcomes and by the probabilities of their occurring; and loudness of tones depends on frequency as well as intensity. In conjoint-measurement theories, the way in which each component affects the attribute in question is studied by discovering which changes must be made in one component to compensate for changes in the other. When measures for the components already exist, the exchange relationship between the two components is often presented graphically in the form of so-called equal-attribute (or indifference) curves.

The simplest representation is additive—or, by making

Order and concatenation

Axioms for quaternary relations

Additive and polynomial representations

an exponential transformation, multiplicative (which is the way it is usually represented in physics)—in the sense that numerical functions ϕ_1 and ϕ_2 of the two components exist such that an empirical inequality holds if and only if the sums of these functions of its elements exhibit a numerical inequality in the same direction; *i.e.*,

$$(a_1, a_2) \geq (b_1, b_2) \equiv \phi_1(a_1) + \phi_2(a_2) \geq \phi_1(b_1) + \phi_2(b_2).$$

Axioms sufficient for such a numerical representation consist of weak ordering, a form of solvability, an Archimedean property, and what amount to analogues of weak monotonicity. The simplest of these, independence, states that an inequality of one component with a common value—*e.g.*, $(a_1, x) \geq (b_1, x)$ —is not altered by substituting any other common value y for x . When this property is appropriately generalized to three or more components, no more stipulations are needed; for only two components, however, it must be supplemented by another property, called the Thomsen condition. In the special case when the components each have only a finite number of elements, a theoretical schema of necessary and sufficient conditions is known. Such additive measures are interval scales; but the multiplicative constant or scale factor is the same for all ϕ_i .

Other ordered representations also have been studied. Especially well understood are certain simple polynomials on several components. For three components there are four such simple polynomials: additive ($\phi_1 + \phi_2 + \phi_3$), multiplicative ($\phi_1\phi_2\phi_3$), or a combination of both, either ($\phi_1\phi_2 + \phi_3$) or $\phi_1(\phi_2 + \phi_3)$. (In these expressions the subscripted indices 1, 2, 3 can, of course, appear in any of their permutations.)

Axioms of geometry. Historically, numerical-representation and uniqueness theorems for geometry have been at least as important as those for extensive measurement and more important than the other kinds of measurement considered above. Since the discovery of analytic geometry in the 17th century, by René Descartes, the earliest important modern philosopher, the representation of geometric points by a pair of numbers (the coordinates), in the case of a plane, or a triple of numbers, in the case of space, has been of fundamental importance not only in geometry but also in much of the physical sciences.

The view that the formal adequacy of axioms for geometry is established by proving a representation theorem became explicit in the latter part of the 19th century. The most important single work may have been the classic *Grundlagen der Geometrie* (1899; *The Foundations of Geometry*, 1902), by David Hilbert. Today the distinctive feature of geometric measurement is that the representation is made in terms of pairs or triples or, in some applications in the social sciences, n -tuples of numbers, rather than in terms of simple numbers, as in extensive, difference, or conjoint measurement.

Much of the literature on the foundations of geometry has been devoted to showing the different sorts of qualitative concepts that can be taken as basic or primitive in stating the axioms of geometry. One of the simplest sets of axioms, given by Alfred Tarski, a Polish-U.S. mathematician and logician, employs the concept of equidistance and that of betweenness for three points that lie on a line. Many of the properties of the quaternary relation of absolute difference discussed above (see above *Axioms of difference*) are also properties of equidistance viewed as a quaternary relation among four points, *viz.*, $(a, b) \sim (c, d)$ just when the distance between a and b is equivalent to that between c and d .

GENERAL PRINCIPLES AND PROBLEMS OF MEASUREMENT THEORY

Dimensions and units of measurement. Each measurable attribute constitutes a dimension; inasmuch as these are not all independent, some can be expressed as (power) functions of others—a fact that underlies both the method of dimensional analysis and the existence of coherent sets of units.

Dimensions and their algebra. Many attributes are extensively measurable. Some dimensions, however, such as density, are not extensive; and in such cases a law can be

stipulated that allows the derived, nonextensive measure to be expressed as a product of powers of two extensive measures. This possibility arises from two empirical facts. First, mass varies both with substance and volume, and the ordering that it induces over substance-volume pairs is such that a multiplicative conjoint representation exists. Thus, the conjoint measure of substance can be expressed as a product of its measure of mass and the reciprocal of its measure of volume. Both of these attributes, of course, have independent extensive measures. Second, a qualitative law, known as a law of similitude, relates concatenations of volume to those of mass via the conjoint ordering. From this law, it is possible to prove that the conjoint and extensive measures are power functions of each other; indeed, in this case, the conjoint measure of substance is, for an appropriate choice of exponent, simply the ratio of the extensive measure of mass to that of volume; *i.e.*, density. In cases such as the dependency of kinetic energy and momentum on mass and velocity, the conjoint components are both extensive and the law relating them via the conjoint ordering is called a law of exchange; interesting cases of such laws appear to involve quantities that enter laws of conservation, such as that of momentum and energy. In any event, all nonextensive measures of classical physics are expressible as products of powers of extensive ones. Some such model is the usual starting point of dimensional analysis.

Universal, system, and material constants. Sometimes two apparently distinct attributes are not independent but covariant. In the classic example of inertial mass and gravitational mass, the covariation is described by a constant known as the universal gravitational constant. This example should be contrasted with the covariation of two measures for specific systems. Mass and volume, for example, covary perfectly for any homogeneous substance, or length and force (within limits) for a specific spring. The constants describing such local covariation, called material or system constants (depending on the context), constitute a form of derived measurement.

Many more complex physical laws may be described as stating combinations of values—configurations—of certain dimensions that can obtain in a particular class of physical system. Such laws include not only measurable dimensions of the system but also system and material constants characteristic of the particular system in question. A curious and not fully understood fact of physical theory is a principle first explicitly enunciated in 1914 by the U.S. analyst Edgar Buckingham, the so-called π -theorem, according to which, when the dimensional measures and constants of some physical law are grouped into one or more terms in which the dimensions all cancel out, some function of these dimensionless quantities (or π -arguments) must equal zero for any realizable configuration of the system.

Dimensional analysis. In dimensional analysis, the investigator in part attempts to discover a particular physical law by assuming that it has the character of a π -theorem and that the relevant variables and constants are known. When they are known (an important proviso), a simple calculus permits him to discover all of the dimensionless terms. In many cases, this yields considerable insight into the law describing the system; sometimes empirical observations, as in a wind tunnel, are used to obtain an empirical approximation to the unknown function of the π -arguments.

Primary and derived units. One is free to choose the unit of measurement for each dimension, whereupon all other values are uniquely determined. Since some dimensions are related to others as products of powers, considerable simplicity is effected by choosing arbitrarily only the units of the dimensions in a maximum set of independent dimensions (base) and then letting the known dependencies determine all of the other units. Such a system of units is called coherent; those of the base are said to be primary, and all others are derived, or secondary. At the Tenth General Conference on Weights and Measures (1954), length, mass, time duration, temperature, and either charge or current were adopted as the base dimensions, with units: metre, kilogram, second,

Units of measurement

Dimensions and constants

degree absolute (Kelvin), and the coulomb or ampere, respectively. This system is abbreviated MKSA.

For effective scientific communication, it is essential that each unit be specified and reproducible to some known degree of accuracy. The ideal definition of a unit is in terms of some highly invariant and readily reproduced or observable natural phenomenon, such as the wavelength of a highly reproducible, monochromatic light source. Less ideal, although still used, are carefully made and maintained unique objects, such as the standard metre in Paris, which have the drawback of being in danger of damage or deterioration.

Measurement via nonmeasurement theories. Few actual measurements are carried out in a fundamental manner using only ordering and concatenation information, because such procedures, when possible, are too slow and expensive, and in fact they are not always possible. It will be helpful to describe some of the alternative procedures employed in physics and psychology.

Indirect
measurement

In physics. At least three kinds of indirect measurement occur in physics. First, permanent standard sequences are sometimes constructed—as for weight and length—and measurement then involves approximate matching against members of the sequence. Precision and standardization of conditions are important in making such measurements.

Second, when a law is found to hold between two or more variables, and system or material constants enter—such as the density and viscosity of a fluid—these constants are measured indirectly in terms of other measures by means of the law.

Third, various theories of specific physical phenomena may provide indirect means of measuring an attribute that may otherwise be difficult to measure; and such theories may, in some cases, lead to convenient measuring devices. A mechanical displacement in a piece of equipment—a spring balance, for example, or a voltmeter—may be systematically related to some other attribute (force or potential). In such measurement, considerable care must be taken to calibrate the apparatus in terms of fundamental measures of the two attributes. Another sort of example is provided by the measurement of astronomical distances in terms of the shift of the object's spectral lines toward the red, a technique that is based on a theory of the existence of a systematic relation between the velocity of recession and the distance of a galaxy, and one about the Doppler effect, which spreads out the waves from a receding light source and thus reduces their frequency.

In psychology. Since few, if any, psychological attributes are currently measured fundamentally, no approximate measurement using standard sequences is possible. Whether conjoint measurement will alter this situation is not yet clear.

Most formalized theory in psychology is stated in terms of physical characteristics of the stimuli and responses: various physical and temporal aspects of the stimuli, the physical nature and probability structure of the rewards provided for appropriate performance, and the time and nature of the response made. Frequently, some of the data are collapsed into relative frequencies, which are interpreted as estimates of conditional probabilities. Though, in principle, a theory could simply assert the dependence of response measures upon the various physical measures of the independent variables, this has not proved practicable; and in most theories some sort of internal representation of the stimuli is postulated to exist in the subject. Decision processes are assumed to operate on these representations, which decisions then lead to the response. Such models include free parameters that often have natural interpretations as psychological measures. If these can be estimated from the observable data, as is frequently possible, they constitute a type of indirect psychological measurement not unlike the second class of physical ones. In many cases, the parameters seem to play a role much like that of time constants in physical systems.

There is no satisfactory analogue to the third method enumerated above for physics, because no sufficiently

elaborate and accepted psychological theories yet exist. Many so-called operational measures of psychological concepts—hours of deprivation for hunger or skin resistance for anxiety—pretend to be of this type, but they are not. The absence of questions of calibration is but one indication of the inadequacy of such measures.

The problem of error. Inasmuch as measurement, in practice, is always fraught with error, the types and basic theory of error must be examined.

Types of error. The classification of errors used in the theory of observations has been of great importance, both historically and practically, especially in astronomy. In that theory, four kinds of errors of observation are ordinarily distinguished. There are instrumental errors, such as those of the telescope in astronomy, that have as their source inaccuracies in the instruments used for observation. There are personal errors, sometimes called errors due to the personal equation, that arise from the different practices and reactions of human observers. Nevil Maskelyne, the fifth astronomer royal of Great Britain, provided a famous example of such personal differences in observation when he discharged his assistant in 1796 because he had observed the transits of stars and planets about a half a second later than Maskelyne himself. Systematic errors are a third kind of error of observation. A typical example of this kind would be the measurement of atmospheric pressures corrected to sea level in terms of a faulty value for the height of the station. Such an error would introduce a systematic discrepancy in all observations of atmospheric pressure for that station. The fourth and most important kind of error of observation is that of random error, so called because its causes are not understood. This type of error reflects the fact that, when observers make repeated measurements, some variation in the results will occur no matter how accurate the instruments may be.

In principle, all four types of error need to be taken into account in refining observations. In practice, however, the attention of theorists has been directed mainly to random errors. Since the end of the 18th century, the application of probabilistic methods through the estimation of random errors has been extensive.

Next to errors of observation, perhaps the most important kind of error is that due to sampling. If an investigator, for instance, wishes to know something about the television-viewing habits of the population, he ordinarily cannot plan to observe the viewing habits of each and every member of the population. He selects, instead, a sample of the population and from that sample infers the viewing habits of all. Errors of sampling, of course, are expected; but, the larger the sample, other things being equal, the smaller the sampling error will be. An elaborate statistical theory, both for constructing the kind of sample that should be drawn and for making inferences from the sample data to the characteristics of the population, is widely used (see STATISTICS).

Still another classification of errors involves the division into direct and indirect errors. If a surveyor, for example, measures the angles of a triangle, direct errors can arise from the measurement of the first two angles. These errors will induce errors of measurement indirectly into the value of the third angle, which he will compute instead of measuring directly. In advanced scientific practice, a large number of errors of measurement are indirect, because elaborate theory and a wide variety of ancillary measurements enter into the measurements of significant quantities, such as the velocity of light or the magnitude of the Earth's mass.

Still other kinds of errors are concerned with interpolation and extrapolation. In using mathematical tables, for example, one often wants to read a table to one more decimal place than those tabulated, and it is common to make a linear interpolation between two adjacent figures in order to extend the table. Since the curve reflected in the figures, however, may in fact be parabolic or exponential or the like, such an assumption of linearity may generate error. Similar problems arise in extrapolating beyond the limits of a given set of values. In both cases it is possible to compute the approximate error introduced.

Errors of
observa-
tion

Other
kinds of
error

Standard error and variance

Theory of errors. The theory of errors has received a great deal of attention in the development of science and mathematics, especially in that of probability theory. Though, historically, the concept of standard error has been most widely used in reporting the results of measurement, in modern statistics the concept is of less importance. The computation of the standard error is based on the assumption that repeated observations of the same phenomenon—for example, the velocity of light—are to be assigned equal weight. It is useful to explore a case of n observations of a phenomenon, x_1, \dots, x_n . If \bar{x} is the mean of these observations—*i.e.*, $\bar{x} = \frac{1}{n}(x_1 + x_2 + \dots + x_n)$ —and s is the sample standard deviation, defined as the square root of the mean squared deviation—*i.e.*,

$$s = \sqrt{\frac{1}{n}[(x_1 - \bar{x})^2 + \dots + (x_n - \bar{x})^2]},$$

then the standard error of \bar{x} is s/\sqrt{n} , and it is said that the “true” value of the observed variable lies in the interval between \bar{x} plus s/\sqrt{n} and \bar{x} minus s/\sqrt{n} . A more detailed interpretation of the standard error and a critical scrutiny of the assumptions underlying its use are beyond the scope of this article (see STATISTICS; also see below *Bibliography*, the work by I. Todhunter). A few brief remarks will have to suffice here.

Under very general assumptions, it may be shown that the sum of independent random errors of observation has, as the number of terms becomes large, a normal (bell-shaped) distribution with mean μ and standard deviation σ , in which μ is the sum of the means of the individual sources of error and σ^2 is the sum of their individual variances (standard deviations squared); *i.e.*, $\sigma^2 = \sigma_1^2 + \sigma_2^2 + \dots + \sigma_q^2$, in which there are q sources of error. (For discussion of the normal distribution with population mean μ and variance σ , see STATISTICS.) This formulation, known as the central limit theorem, is one of the fundamental theorems of modern probability theory. Its importance lies in the fact that it is largely independent of the particular character or distribution of the individual errors. It was first stated by Laplace in 1812 and given its first rigorous proof in 1901 under fairly general assumptions by the Russian Aleksandr Lyapunov, known for his mathematical studies of stability and probability.

Probabilistic models

In recent years different probabilistic models have been developed for the study of human behaviour, especially behaviour that expresses a choice or preference among several alternatives. Such models are devised to measure the strength of preference so as to predict the probability of choice—for experience in everyday life and in the laboratory has shown that men and animals will vary their choice of one alternative over another in what are externally identical conditions. One simple and useful model is this: if $v(x)$ is defined as the utility or value of x —*i.e.*, the numerical measure of the strength of preference for x —and $p(x,y)$ as the probability of choosing x over y , then the model postulates that $p(x,y) = v(x)/[v(x) + v(y)]$. The problem for the theory of measurement is to show what properties the observed probabilities $p(x,y)$ must have in order for such a utility function v to exist. If a child, for instance, chooses chocolate ice cream (c) over licorice (l) with probability two-thirds, licorice over a banana (b) with probability two-thirds, and b over c with probability two-thirds, then no such strength of preference function v can be found for the set $\{b,c,l\}$.

MEASUREMENT IN SPECIAL DISCIPLINES

Classical physics

Measurement in physics. The development of the central concepts of classical physics—such as the mechanical concepts of mass, force, momentum, and kinetic energy, and electromagnetical concepts such as those of current, voltage, resistance, and impedance—has been closely intertwined with the development of procedures for measuring quantitatively the properties of the phenomena associated with the concept. The theory and practice of extensive measurement, as discussed above (see above

Axiomatic basis of measurement), and the theory of dimensions and units (see above *Dimensions and units of measurement*) have been developed mainly in the context of classical physics.

One of the most important attitudes toward measurement in classical physics was the firmly rooted belief that, with sufficient effort, errors of measurement could be eliminated in principle and that there would be no limitation to the precision of measurement and, consequently, to the accuracy to which theories might be pushed in testing their validity. Although classical physics continues to be of major importance in applied science, a much more restrained attitude toward measurement now is customary. In the applications of classical electromagnetic theory, for example, it is customary to impose a lower limit defining the smallest admissible length to be considered: a typical restriction is to keep all lengths greater than one-tenth of a millimetre in order to avoid any atomic or subatomic effects that would fall properly within quantum theory.

Modern physics

A central concept of Einstein's theory of relativity and one that has had an enormous impact on modern physics is its revision of the classical concept of simultaneity, which restricts it to immediately adjacent events. There is no well-defined concept of two events at astronomical distances occurring simultaneously; this means that classical procedures for the measurement of simultaneity have been shown to be mistaken in their fundamental assumptions and that the concept of simultaneous distant events is as mistaken as that of absolute position in space.

At another level of analysis are the so-called Lorentz transformations of special relativity, which formulate the relations between two differently moving space-time frames. It is possible to give a self-contained and elementary derivation of these transformations (without any assumptions of continuity or linearity) from a single axiom asserting that the relativistic interval between any two space-time points connected by a possible inertial path of a particle remains the same in all inertial frames of reference. In other words, just two stipulations—that the macroscopic measurement of such relativistic distances be conducted along inertial paths (those followed naturally by unaccelerated particles) and that these measurements be invariant for any inertial frame of reference—are sufficient to establish the fundamental principle that any two inertial frames of reference are related by Lorentz transformations.

The theory of measurement has been important in quantum mechanics for two distinct but related conceptual reasons. One deals with the fundamental discovery that it is not possible to measure simultaneously with arbitrary precision the position and momentum of subatomic particles such as, for example, electrons or photons. The best known way of expressing this fact is in terms of the Heisenberg uncertainty principle, so-called after Werner Heisenberg, a German quantum physicist, which states that the product of the standard error of measurement of position and momentum must exceed a certain positive constant (for a more detailed discussion, see MECHANICS, QUANTUM).

The strangeness of this discovery, from the standpoint of classical physics, has been one of the most discussed problems of modern physics. Efforts to account for the theoretical uncertainties in measurement or the theoretical limits on the precision of measurement have been a central focus of conceptual discussions of the foundations of quantum mechanics. From the standpoint of classical statistical theory, perhaps what is most surprising about the Heisenberg uncertainty principle and related aspects of quantum mechanics is that one is not able to study the covariance of position and velocity measurements. The necessity of looking only at the marginal probability distributions of position and momentum and at the marginal distributions of other conjugate observables in quantum mechanics is the source of the sharpest break between the methodology of quantum mechanics and that of classical statistical theory as used both in classical physics and in other domains of science.

The second aspect of measurement that has been funda-

mental in quantum mechanics is the interaction between the measurement instrument and the measured object. In classical physics it was assumed that the interference of the measuring instrument can in principle be reduced to zero; in quantum mechanics the complicated and detailed analysis of the situation, in which the opposite is a fundamental assumption, has proved to be difficult and elusive.

Measurement in the social sciences and psychology. These fields pose special problems for measurement theory.

Economics and psychology. An important reason why economics is the most advanced of the social sciences—advanced in predictive power, social usefulness, and formal theory—is that the many important economic variables may be easily measured. Much of both economic theory and data are cast in terms of price and quantity. The problems associated with these concepts are not so much ones of measurement as they are of aggregation to provide representative measures for large economic units, such as an entire industry or a whole economy. Social indices, such as cost of living or gross national product, involve subtle questions of equivalence, statistical sampling, and the like.

In spite of their success in using the concepts of price and quantity, economists recognize that individuals often base their decisions on more subjective concepts. One approach is to suppose that such individuals have a numerical measure of utility, which they try to maximize. Early attempts to introduce utility into economics employed a representation that, in essence, assumed that the utility of a bundle of commodities is the sum of the utilities of the several commodities; but little attempt was made to axiomatize the representation, and, for various reasons, that hypothesis was discredited. In the 1940s it was suggested that utility could be measured by taking into account the risky character of decision alternatives. Taking a and b as the alternatives and P as a probability, these scholars symbolized by aPb the risky alternative in which one receives either a or b , the former with probability P and the latter with probability $1 - P$. They then axiomatized orderings over such structures to show that an order-preserving numerical assignment exists with the expected-utility property: $u(aPb) = u(a)P + u(b)(1 - P)$. Under these conditions, decision making becomes equivalent to maximizing expected utility. This idea has received considerable development since the end of World War II. Its main thrust has been the more careful formulation of what an alternative is in terms of various outcomes associated with different chance events and the generalization of the representation theorem to one in which, in addition to constructing the utility function, a (subjective) probability function over events is also constructed from the preference ordering of the alternatives in such a way that subjective expected utility is order preserving.

Although these ideas originated in economics and statistics, psychologists were quick to recognize the problem as one of individual preferences, and they have run numerous experiments to assess its descriptive adequacy. To oversimplify considerably, the major qualitative property known as the (extended) sure-thing principle is not valid. Roughly, this property asserts that if, no matter what possibility is realized, the outcomes of one gamble are at least as satisfactory as the corresponding ones of another, then the former gamble will be preferred or indifferent to the latter one—in the above notation, $aPb \geq a'P'b' \equiv (aPb)Qc \geq (a'P'b')Qc$. This principle is analogous to independence in additive conjoint measurement and to weak monotonicity in extensive measurement. As a postulate of rational behaviour, however, it has not been seriously questioned; thus, interest in the theory remains.

As was noted above (see above *In psychology*), some psychological theories include parameters that may be interpreted as measures of psychological attributes, parameters of which prototypic examples are found in the study of the detection and recognition of simple sensory signals. Perhaps the best known example is the theory of signal detectability, adapted from electrical engineering, which has two main parameters: d' , which in effect de-

scribes how detectable the signal is against its background noise, and β , which describes how willing the subject is to interpret evidence as favouring the hypothesis that the signal is present. The former is affected by the signal-to-noise ratio, whereas the latter is manipulated by payoffs and the proportion of trials on which signals are presented. Given these two numbers, the theory specifies how to calculate the response probabilities for a variety of experimental designs.

Another type of psychological research attempts to treat the person as a null-measuring instrument—*i.e.*, as a sensor that records the point at which a difference becomes imperceptible—by having him establish equivalent intensity ratios in different sense modalities, such as sight and sound. These procedures of cross-modality matching yield approximate power relations among the usual physical measures of intensity. Various attempts are being made to place these results in either of two contexts: fundamental measurement theory or elaborations of existing theories of local psychophysics.

Sociology and demography. The main object of sociology as a science is the study of social groups and institutions and the causes of change within them. Many traditional theories in sociology have been of a speculative sort, not easily subjected to quantitative test. An increasing emphasis on quantitative theories and concepts, however, has inspired a greater concern with problems of measurement. The one part of sociology that from its beginning has emphasized quantitative data and analysis is demography, which is concerned with the quantitative study of human populations and the causes of changes in population. Though demographers are interested mainly in the measurement of factors of population change—including births, deaths, and migrations—they are also concerned with questions of population composition in terms of age, sex, and other characteristics. As in the case of sociology in general, there have been few fundamental problems of measurement of a direct sort in demography; the problems have been primarily of a statistical sort—to infer, for example, future birth trends from intensive field studies of selected samples of a population, although the trend is toward greater accuracy. Perhaps the most successful application of measurement⁸ in demography has been in the construction of the mortality tables used by insurance companies and government agencies.

BIBLIOGRAPHY. D.H. KRANTZ *et al.*, *Foundations of Measurement*, vol. 1 (1971), a comprehensive treatise in which the whole axiomatic view, including many examples of representation and uniqueness theorems, is developed extensively; JOHANN PFANZAGL, *Theory of Measurement* (1968), an earlier, briefer general treatise; R.D. LUCE and P. SUPPES, "Preference, Utility, and Subjective Probability," in R.D. LUCE, R.R. BUSH, and E. GALANTER (eds.), *Handbook of Mathematical Psychology*, vol. 3, pp. 249–410 (1965), an extensive review of preference models; P.C. FISHBURN, *Utility Theory for Decision Making* (1970), a concise mathematical summary of results in utility theory; BRUNO KISCH, *Scales and Weights: A Historical Outline* (1965), an excellent survey of the more practical aspects, with special emphasis on the history of the instruments used for scales and weights; DAVID HILBERT, *Die Grundlagen der Geometrie* (1899; Eng. trans., *The Foundations of Geometry*, 1902, reprinted 1950), one of the first fully satisfactory axiomatizations of geometry from a formal standpoint; JOHANN VON NEUMANN and OSKAR MORGENSTERN, *Theory of Games and Economic Behavior*, 3rd ed. (1953), the original treatise on the theory of games; S.S. STEVENS, "On the Theory of Scales of Measurement," *Science*, 103:677–680 (1946), his original classification of scales; ISAAC TODHUNTER, *A History of the Mathematical Theory of Probability from the Time of Pascal to That of Laplace* (1865, reprinted 1949), includes an excellent discussion of the early history of the concept of standard error; JOHANN VON NEUMANN, *Mathematische Grundlagen der Quantenmechanik* (1943; Eng. trans., *Mathematical Foundations of Quantum Mechanics*, 1955), a classical treatise containing what remains one of the best discussions of the effects of instrumentation on measurement.

(R.D.L./P.Su.)

Meat and Meat-Packing

Meat, commonly defined as the edible portion and associated parts of mammals, is usually restricted to beef, veal,

Measurements of price, quantity, and utility

Recent efforts at quantification

Sensation