Patrick Suppes. Theory of Measurement. E. Craig (Ed.), Routledge Encyclopedia of Philosophy. London: Routledge, 1998, pp. 243-249.

MEASUREMENT, THEORY OF

A conceptual analysis of measurement can properly begin by formulating the two fundamental problems of any measurement procedure. The first problem is that of representation, justifying the assignment of numbers to objects or phenomena. We cannot literally take a number in our hands and 'apply' it to a physical object. What we can show is that the structure of a set of phenomena under certain empirical operations and relations is the same as the structure of some set of numbers under corresponding arithmetical operations and relations. Solution of the representation problem for a theory of measurement does not completely lay bare the structure of the theory, for there is often a formal difference between the kind of assignment of numbers arising from different procedures of measurement. This is the second fundamental problem, determining the scale type of a given procedure.

Counting is an example of an absolute scale. The number of members of a given collection of objects is determined uniquely. In contrast, the measurement of mass or weight is an example of a ratio scale. An empirical procedure for measuring mass does not determine the unit of mass. The measurement of temperature is an example of an interval scale. The empirical procedure of measuring temperature by use of a thermometer determines neither a unit nor an origin. In this sort of measurement the ratio of any two intervals is independent of the unit and zero point of measurement.

Still another type of scale is one which is arbitrary except for order. Moh's hardness scale, according to

which minerals are ranked in regard to hardness as determined by a scratch test, and the Beaufort wind scale, whereby the strength of a wind is classified as calm, light air, light breeze, and so on, are examples of ordinal scales.

A distinction is made between those scales of measurement which are fundamental and those which are derived. A derived scale presupposes and uses the numerical results of at least one other scale. In contrast, a fundamental scale does not depend on others.

Another common distinction is that between extensive and intensive quantities or scales. For extensive quantities like mass or distance an empirical operation of combination can be given which has the structural properties of the numerical operation of addition. Intensive quantities do not have such an operation; typical examples are temperature and cardinal utility.

A widespread complaint about this classical foundation of measurement is that it takes too little account of the analysis of variability in the quantity measured. One important source is systematic variability in the empirical properties of the object being measured. Another source lies not in the object but in the procedures of measurement being used. There are also random errors which can arise from variability in the object, the procedures or the conditions surrounding the observations.

1 First fundamental problem: representation

Since the days of the Greek geometers, mathematicians, philosophers and empirical scientists have been discussing the nature of measurement. Eudoxus' development of the theory of proportion, which forms the substance of Book V of Euclid's *Elements*, is probably the first sustained analysis directly relevant to measurement. A systematic approach to the subject of measurement may well begin by formulating the two fundamental problems of any procedure of measurement. The first problem is to justify the assignment of numbers to objects or phenomena.

The early history of mathematics shows how difficult it was to divorce arithmetic from particular empirical structures. Thus the early Egyptians could not think of $2+3$, but only of 2 bushels of wheat plus 3 bushels of wheat. From a logical standpoint, there is just one arithmetic of numbers, not an arithmetic for bushels of wheat, and a separate arithmetic for quarts of milk. The first problem for a theory of measurement is to show how this one arithmetic of numbers may be applied in a variety of empirical situations. This is done by showing that the relevant structure of the empirical situation investigated is isomorphic to an arithmetical structure. The purpose of the definitions of isomorphism given for each kind of measurement considered is to make the rough-and-ready intuitive idea of same structure precise. The great significance of finding such an isomorphism of structures is that we may then use all our familiar knowledge of computational methods, as applied to the arithmetical structure, to infer facts about the isomorphic empirical structure.

A procedure of measurement is needed in any area of science when we desire to pass from simple qualitative observations to the quantitative observations necessary for the precise prediction or control of phenomena. To justify this transition we need an algebra of empirically realizable operations and relations which can be shown to be isomorphic to an appropriately chosen numerical algebra. Satisfying this requirement is the fundamental problem of measurement representation.

2 Second fundamental problem: uniqueness

Solution of the representation problem for a theory of measurement is not enough. There is often a formal difference between the kind of assignment of numbers arising from different procedures of measurement. Consider the following three statements:

(1) The number of people now in this room is seven.

(2) Yesterday John weighed 150 lbs.

(3) The maximum temperature tomorrow will be 89°F.

Here we may formally distinguish three kinds of measurement. Counting is an example of an 'absolute' scale. The number of members of a given collection of objects is determined uniquely. There is no arbitrary choice of a unit or zero available. In contrast, the measurement of mass or weight is an example of a 'ratio' scale. Empirical procedures for measuring mass do not determine the unit of mass. The choice of a unit is an empirically arbitrary decision, but the ratio of any two masses is independent of such choice. The measurement of distance is a second example of a ratio scale. The ratio of the distance between Palo Alto and San Francisco to the distance between Washington and New York is the same whether the measurement is made in miles or metres.

The measurement of temperature is an example of an interval scale. The empirical procedure of measuring temperature by use of a thermometer determines neither a unit nor an origin. In this sort of measurement the ratio of any two intervals is independent of the unit and zero point of measurement. Examples other than measurement of temperature are measurements of temporal dates, geographical longitude and cardinal utility.

In terms of the notion of absolute, ratio and interval scales we may formulate the second fundamental problem for a theory of measurement: determine the scale type of the measurements resulting from the procedure. (A detailed treatment of the theory of fundamental measurement from the standpoint of these two central problems is to be found in Krantz *et al.* the three-volume treatise *Foundations of Measurement*, 1971, 1989, 1990.)

3 Classification of scales of measurement

For the purpose of systematizing some of the discussion of the uniqueness problem in the previous section, we may define a scale as a class of measurement procedures having the same transformation properties. Examples of three different scales have already been mentioned, namely, counting as an absolute scale, measurement of mass as a ratio scale and measurement of temperature as an interval scale.

An 'ordinal' scale is one which is arbitrary except for order. A number is assigned to give a ranking. The social sciences extensively use such scales. Numbers are also sometimes used for naming or classification. The numerical assignment is completely arbitrary. The numbers on football players' shirts are examples of this sort of measurement. Such scales are usually called 'nominal' scales.

In the literature a distinction is often made between those scales of measurement which are fundamental and those which are 'derived'. A derived scale is one which in the procedure of measurement presupposes and uses the numerical results of at least one other scale. In contrast, a 'fundamental' scale does not depend on other scales. As might be expected, foundational investigations of measurement have been primarily concerned with fundamental scales, and such is the emphasis here (Krantz *et al.* 1971: ch. 10).

Another distinction common in the literature is that between extensive and intensive quantities or scales. For extensive quantities an empirical operation of combination can be given which has the structural properties of the numerical operation of addition. Intensive quantities are characterized by the absence of such an operation.

There is a fairly large literature on the problem of characterizing in what sense fundamental measurement of intensive quantities is possible. An early, significant example is the analysis of Nicole Oresme in the fourteenth century.

4 Measurement and theory construction

It has become a platitude of scientific method that measurement of some attribute is really only interesting if some theory about that attribute is available. It is common to criticize many psychometric scales in psychology on the ground that application of the scales leads nowhere. Put another way, the criticism is that measurement for measurement's sake is not very fruitful scientifically.

There are two other aspects of the relation between measurement and theory construction that are important and yet are somewhat neglected. The first is that a theory of measurement for some characteristic of empirical phenomena is itself a genuine scientific theory. Moreover, certain theories are practically coextensive with the theory of measurement of the quantities basic to the theory. For example, the heart of the theory of rational behaviour in situations of limited information is the measurement of subjective probability and utility (see Decision and game theory §2).

The second aspect is that the chief stumbling block to applying certain theories is a problem of measurement. A classical example in physics is the general theory of relativity. Of the three 'traditional' pieces of evidence for the general theory, namely, the advancement of the perihelion of Mercury, the deflection of light waves by the gravitational field of the sun, and the red shift of stellar spectra, only the first does not present really difficult measurement problems (see Einstein, A. §3).

5 Ordinal theory of measurement

As a simple example of fundamental measurement, we may consider the ordinal theory of measurement. Models of this theory are customarily called weak orderings, defined as follows. Let A be a nonempty set and R a binary relation on this set. A pair (A, R) is a 'simple relation structure'. A simple relation structure (A, R) is a 'weak ordering' if and only if for every x, y and z in A (i) if xRy and yRx then xRz, and (ii) xRy or yRx.

The definition of isomorphism of simple relation structures illustrates nicely the general concept. A simple relation structure (A, R) is 'isomorphic' to a simple relation structure (A', R') if and only if there is a function f such that (i) the domain of f is A and the range of f is A', (ii) f is a one–one function, and (iii) if x and y are in A then xRy if and only if $f(x)R'f(y)$. To illustrate this definition of isomorphism let us consider the question: 'are any two finite weak orderings with the same number of elements isomorphic?' Intuitively it is clear that the answer is negative, because in one of the weak orderings all the objects can stand in the relation R to each other and not so in the other.

Homomorphism of models. In many cases a representation theorem in terms of isomorphism of models turns out to be less interesting than a representation theorem in terms of the weaker notion of homomorphism. Good examples are provided by theories of measurement. As already noted, when we consider general practices of measurement it is evident that in terms of the structural notion of isomorphism we would, roughly speaking, think of the isomorphism as being established between an empirical model of the theory of measurement and a numerical model. However, a slightly more detailed examination of the question indicates that difficulties about isomorphism quickly arise. In all too many cases of measurement, distinct empirical objects or phenomena are assigned the same number, and thus the one–one relationship required for isomorphism of models is destroyed. Fortunately, this weakening of the one–one requirement for isomorphism is the only respect in which we must change the general notion, in order to obtain an adequate account for theories of measurement of the relation between empirical and numerical models. The general notion of homomorphism is designed to accommodate exactly this situation. To obtain the formal definition of homomorphism for two simple relation structures as previously defined, we need only drop the requirement that the function establishing the isomorphism be one–one.

The concept of homomorphism is what we need to state a representation theorem for weak orders. First, a 'numerical' weak ordering is a weak ordering $(A, \leq)$ where A is a set of numbers. The selection of the numerical relation $\leq$ to represent the relation R in a weak ordering is arbitrary, in the sense that the numerical relation $\geq$ could just as well have been chosen. However, choice of one of the two relations $\leq$ or $\geq$ is the only intuitively sound possibility. The following theorem provides a homomorphic representation theorem for finite weak orderings, and thus makes the theory of finite weak orderings a theory of measurement:

(1) Every finite weak ordering is homomorphic to a numerical weak ordering.

(1) was restricted to finite weak orderings for good reason: it is false if this restriction is removed. In order to state the desired theorem for the infinite case one preliminary notion is needed. Let (A, R) be a simple relation structure, and let B be a subset of A. Define the strict ordering relation P in terms of R by the equivalence: xPy if and only if xRy and not yRx. Then we say that B is *R-order-dense* in A if and only if for every x and y in A and not in B such that xPy there is a z in B such that xRz and zRy. Note that the denumerable set of rational numbers is order-dense in the non-denumerable set of all real numbers with respect to the natural numerical ordering $\leq$. This relationship between the denumerable rational numbers and all real numbers is just the one that is necessary and sufficient for the existence of a homomorphism between an infinite and a numerical weak ordering.

(2) Let (A, R) be an infinite weak ordering. Then a necessary and sufficient condition that it be homomorphic to a numerical weak ordering is that there is a denumerable subset B of A such that (i) B is R-order-dense in A and (ii) no two elements of B stand in the relation E, for example, for any distinct x and y in B either not xRy or not yRx.

The proof is related to the classical ordinal characterization of the continuum by Cantor (1895) (see CONTINUUM HYPOTHESIS). The formulation and proof of (2) uses more mathematical apparatus than elementary logic. The restriction of the discussion of theoretical issues in the philosophy of science to the framework of first-order logic has too often meant a restriction of the extent to which the discussion could come into contact with other than highly simplified scientific theories especially constructed for the purpose at hand. An advantage of the set-theoretical formulation of theories such as theories of measurement is that standard mathematical methods are immediately available for both the formulation of the theory and the deductive analysis of the structure of its models (see THEORIES, SCIENTIFIC §3).

6 Invariance and meaningfulness

As already mentioned, the number assigned to measure mass is unique once a unit has been chosen. A more technical way of putting this is that the measurement of mass is unique up to a similarity transformation. A numerical function ϕ is a similarity transformation if there is a positive number α such that for every real number x, $\phi(x) = \alpha x$. In transforming from pounds to grams, for instance, the multiplicative factor α is 453.6. The measurement of temperature in °C or F has different characteristics. Here an origin as well as a unit is arbitrarily chosen.

Technically speaking, the measurement of temperature is unique up to a linear transformation. A numerical function ϕ is a 'linear' transformation if there are numbers α and β with $\alpha > 0$ such that for every number x, $\phi(x) = \alpha x + \beta$. In transforming from centigrade to Fahrenheit degrees of temperature, for instance, $\alpha = 9/5$ and $\beta = 32$.

Ordinal measurements are commonly said to be unique up to a monotone increasing transformation. A numerical function ϕ is a monotone increasing transformation if, for any two numbers x and y, if $x < y$, then $\phi(x) < \phi(y)$. Such transformations are also called 'order-preserving'.

An empirical hypothesis, or any statement in fact, which uses numerical quantities is empirically meaningful only if its truth value is invariant under the appropriate transformations of the numerical quantities involved. As an example, suppose a psychologist has an ordinal measure of IQ, and thinks that scores $S(a)$ on a certain new test T have ordinal significance in ranking the intellectual ability of people. Suppose further that the psychologist is able to obtain the ages $A(a)$ of the subjects. The question then is: should the following hypothesis be regarded as empirically meaningful? For any subjects a and b, if $S(a)/A(a) < S(b)/A(b)$, then $\mathrm{IQ}(a) < \mathrm{IQ}(b)$. From the standpoint of the invariance characterization of empirical meaning, the answer is negative. To see this, let $\mathrm{IQ}(a) \geq \mathrm{IQ}(b)$, let $A(a) = 7$, $A(b) = 12$, $S(a) = 3$, $S(b) = 7$. Make no transformations on the IQ data, and make no transformations on the age data. But let ϕ be a monotone-increasing transformation of the ordinal measurements $S(a)$ which carries 3 into 6 and 7 into 8. Then we have $3/7 < 7/12$, but $6/7 \geq 8/12$, and the truth value of the hypothesis is not invariant under ϕ.

The empirically significant thing about the transformation characteristic of a quantity is that it expresses in precise form how unique is the structural isomorphism between the empirical operations used to obtain a given measurement and the corresponding arithmetical operations or relations. If, for example, the empirical operation is simply that of ordering a set of objects according to some characteristic, then the corresponding arithmetical relation is that of less than (or greater than), and any two functions which map the objects into numbers in a manner preserving the empirical ordering are adequate. Only those arithmetical operations and relations which are invariant under monotone transformations have empirical significance in this situation.

We can easily state the uniqueness or invariance theorem corresponding to the representation theorem (1) for finite weak orders:

(3) Let (A, R) be a finite weak order. Then any two numerical weak orderings to which it is homomorphic are related by a monotone-increasing transformation.

Put in other language, the numerical representation of finite weak orders is unique up to an ordinal transformation. Invariance up to ordinal transformations is not a very strong property of a measurement, and it is for this reason that the hypothesis considered above turned out not to be meaningful, because it was not invariant under monotone transformations of the measurement data.

A measurement representation theorem should ordinarily be accompanied by a matching invariance theorem stating the degree to which a representation of a structure is unique. In the mathematically simple and direct cases of measurement it is usually easy to identify the invariant group of transformations. For more complicated structures, for example structures that satisfy the axioms of a scientific theory, it may be necessary to introduce more complicated apparatus, such as the Galilean or Lorentz transformations of physics, but the objective is the same, namely, to characterize meaningful concepts in terms of invariance.

One point needs emphasis. When the concepts are given in terms of the representation, for example a numerical representation in the case of measurement, or representation in terms of Cartesian coordinates in the case of geometry, a test for invariance is needed. When purely qualitative relations are given which are defined in terms of the qualitative primitives of a theory, for example those of measurement or geometry, then it follows at once that the defined relations are invariant and therefore meaningful. On the other hand, the great importance of the representations and the reduction in computations and notation they achieve, as well as understanding of structure, make it imperative that we have a clear understanding of invariance and meaningfulness for representations which may be, in appearance, rather far removed from the qualitative structures that constitute models of the theory (Krantz *et al.* 1990: ch. 22).

7 Variability, thresholds and errors

In the standard theory of fundamental measurement as discussed, qualitative axioms are formulated that lead to a numerical assignment unique up to some simple transformation. A widespread complaint about this classical foundation of measurement is that it takes too little account of the analysis of variability in the quantity measured. The sources of such variability can be of several different kinds. One important source is simply variability in the empirical properties of the object being measured. The height of a person for example varies on a diurnal basis and the variability of the weight of a person from day to day is a familiar fact of everyday experience. A quite different source of variability lies not in the object being measured but in the procedures of measurement being used. When the procedures being used lead to variability, these are ordinarily attributed to error. The study of error, especially in physical measurements goes back to the eighteenth century with fundamental work by Simpson, Lagrange, Laplace and, especially, Gauss. In the physical sciences the standard reporting of statistical errors in scientific publications began about the middle of the nineteenth century, especially in the publication of astronomical observations.

At least five kinds of error have been distinguished in astronomical observation, and they illustrate nicely the kind of analysis that can be made of errors in other empirical observations, even though changes in focus must be made when the empirical phenomena are quite different. For astronomers there are instrumental errors arising from imperfections in the production of the instruments of observation used such as telescopes and clocks. There are personal errors due to the response characteristics of the observer, for instance, coordinating a visual observation with the auditory beat of a clock, or with the nearly simultaneous visual observation of a clock, will lead to slightly different results for different observers. One example, Nevil Maskelyne, the fifth astronomer royal of Great Britain, discharged his assistant in 1796 because the assistant had observed the transits of stars and planets about a half-second later than Maskelyne himself. An objective of modern technology is to eliminate as much as possible such personal errors, but they certainly persist in many areas of empirical observation. Errors of the third kind are 'systematic' and are due to conditions that are themselves subject to observation and measurement. For example, steady winds in a given direction must be taken account of in many kinds of familiar procedures, such as a navigator setting a course of a ship or aeroplane. Fourth, there are 'random' errors which can arise from variability in the conditions surrounding the observations and are not necessarily due either to variations in the object being measured or in the procedures of measurement. Meteorological variations affecting astronomical observations are a good example. Of a different sort, and a fifth kind, are empirical errors of 'computation' once numerical observations have been recorded. It is important to emphasize that there is nothing fixed and a priori about the classification given here. Other classifications can be found in the literature, but there is a natural conceptual basis to support each of the types of error listed (Taylor 1982).

It is not possible here to examine in detail how the foundational investigations of measurement procedures have been able to deal with such problems of error. One point of interest has been the analysis of thresholds in psychological judgments. It is a familiar fact that if a person is asked to judge the weight of two objects that are very close together, psychological discrimination of a difference will not occur when the difference is sufficiently small. There is a large body of work and a large body of theory connected with such psychological thresholds (Krantz *et al.* 1989: ch. 16).

A central objective of foundational analysis in this area is to begin with qualitative axioms, but to expect the representation for a measurement property of an object to be in terms of a random variable with a given probability distribution rather than in terms of a fixed numerical assignment. Such assignment of random variables rather than numbers to objects is already an abstraction from the various kinds of measurement error distinguished above. Part of the reason for this is that the main objective of such analysis is to end up with probability distributions for the errors in the form of some standard probability distributions familiar in mathematical statistics and amenable to detailed statistical treatment.

The theory of measurement has been important in quantum mechanics, the most significant physical theory of the twentieth century, for two distinct but related conceptual reasons. One arises from the Heisenberg uncertainty principle which shows that it is not possible to measure simultaneously with arbitrary precision the position and momentum of particles such as electrons or protons. Closely connected with this is the second aspect of measurement in quantum mechanics, namely the recognition that there is a physical interaction between the measurement instrument and the measurement object, a subject not really developed at all in classical physics (von Neumann 1996; see QUANTUM MEASUREMENT PROBLEM).

In the social sciences and medicine a familiar kind of error is that which arises from finite sampling. If we wish to know something about the telephone habits, driving habits or reactions to a new medicine of a given population, we ordinarily do not plan to interview or test everyone in the population, but rather use a random sample to infer the distribution of habits or reactions in the entire population. The theory of such sampling is very much a twentieth-century subject (see STATISTICS AND SOCIAL SCIENCE §2).

See also: EXPERIMENT; OBSERVATION; OPERATIONALISM; THEORY AND OBSERVATION IN SOCIAL SCIENCES

References and further reading

* Cantor, G. (1895) 'Beiträge zur Begründung der transfiniten Mengenlehre', *Mathematische Annalen* 46: 481–512. (See §5.)

Clagett, M. (ed.) (1968) *Nicole Oresme and the Medieval Geometry of Qualities and Motions*, Madison, WI: University of Wisconsin Press. (Contains Oresme's *Tractatus de configurationibus qualitatum et motuum*, with an introduction, translation and commentary.)

Kisch, B. (1965) *Scales and Weights: A Historical Outline*, New Haven, CT: Yale University Press. (A detailed survey of the more practical aspects, with special emphasis on the history of the instruments used for scales and weights.)

* Krantz, D.H., Suppes, P., Luce, R.D. *et al.* (1971, 1989, 1990) *Foundations of Measurement*, vol. I, *Additive and Polynomial Representations*, vol. II, *Geometrical, Threshold, and Probabilistic Representations*, vol. III, *Representation, Axiomatization, and Invariance*, New York: Academic Press. (A comprehensive treatise in which the fundamental problems of measurement are given extensive treatment with many examples of representation and uniqueness theorems, and extensive references to the literature in many disciplines on the theory of measurement.)

* Neumann, J. von (1996) *Mathematical Foundations of Quantum Mechanics*, Princeton, NJ: Princeton University Press. (A classical treatise containing one of the most thorough and important discussions of the interaction between the measurement instrument and the measured object, with a particular reference to quantum mechanics but also applicable to measurement procedures generally.)

* Taylor, J.R. (1982) *An Introduction to Error Analysis*, Mill Valley, CA: University Science Books. (An excellent elementary text on uncertainties in physical measurement.)

PATRICK SUPPES