MEASUREMENT

Measurement is widely performed to produce information about properties of objects, and the quality of measurement results may be evaluated: a theory of measurement provides criteria to distinguish measurements from processes like opinion making, and about the quality of measurement results. Moreover, it establishes the structure – whether quantitative or not only – and the domain – whether physical or not only – of properties to be measurable.

The scope of a theory of measurement

Measurements are performed in many different contexts to produce information about properties of systems, phenomena, processes, individuals, organizations, and so on. It is an integral part of modern science as well as of engineering, commerce, and daily life. Measurement is often considered a hallmark of the scientific enterprise and a privileged source of knowledge. But what is the source of the acknowledged *special efficacy* of measurement?

First, a theory of measurement is expected to provide a criterion of *demarcation*, able to distinguish measurement from other information-related processes like opinion making and computation: not all processes of information production are measurements. Second, the fact that measurement is a designed-on-purpose process, not a natural phenomenon, adds a pragmatic layer to the theory: a theory of measurement is expected to provide some criteria of *qualification*, able to characterize good and bad measurements (the quality of an opinion, for example by a highly qualified expert, could be better than the quality of a measurement, as performed for example by means of a cheap instrument), and this prevents defining measurement simply as any high-quality information production process.

Moreover, a theory of measurement needs to be situated with respect to two strategic issues. First, is there any constraint about the *structure* of measurable properties? And therefore, in particular: is – for some reason to be identified – measurability a feature of quantitative properties ("quantities" for short) only, or can also non-quantitative properties be measured? Second, is there any constraint about the *domain* of measurable properties? And therefore, in particular: is – for some reason to be identified – measurability a feature of physical properties only, or can also psychosocial properties be measured?

These issues probe the very idea of what measurement is, as developed across millennia of scientific, technological, philosophical, and societal changes, and as currently challenged by the widespread emphasis on datafication: can an entirely informational process, with no empirical stages in it, be a measurement (perhaps a "virtual measurement", as sometimes it is called)? Does a difference remain between a measurement and a simulation of a measurement?

After some notes setting a fundamental terminological ground, the domain – whether physical or not only – and the structure – whether quantitative or not only – of measurable properties are analyzed in a historical perspective, and on this basis the problems of demarcation and qualification are discussed. The conclusion will be that what is measurement, and therefore what a theory of measurement should deal with, is still a moving target.

To be published in The SAGE Encyclopedia of Theory in Science, Technology, Engineering, and Mathematics

Preprint

Background

Measurement is about attributing values to properties of objects: unpacking this is a precondition for a theory of measurement. The assumption is that *objects*, like systems, phenomena, processes, individuals, and organizations, *have properties*, like length, color, age, and intelligence (in a property-oriented ontology an object is something that has properties; whether objects can be identified independently of their having properties is an issue preliminary to our subject). Not any object *o* has any property *P* (for example gasses and companies do not have geometric shape), so that for each property the set of objects is partitioned into the subset of the objects that have the property – it may be called the *domain* of the property, *D*(*P*) – and in the complementary subset of the objects that do not have it. This may be written in the predicative form *has_P*(*o*), that is either true or false, so that

$D(P) := \{o: has\_P(o)\}$

Hence *has_electric_charge*(*electron*) is true and *has_reading_comprehension_ability*(*Moon*) is false. Such a classification – for any sufficiently well-identified *P* and *o*, either *has_P*(*o*) or not – is the piece of information that provides the fundamental justification of the interest in dealing with properties.

Sometimes this information may be refined, by recognizing that a property *P* is associated with a set X_P of other properties, $X_{P,1}$, $X_{P,2}$, ... such that, for each *o*, *has_P*(*o*) if and only if there exists one and only one $X_{P,i}$ in X_P such that $has\_X_{P,i}(o)$. For example, for a physical body *o*, *has_color*(*o*) if and only if *has_color_red*(*o*), or *has_color_blue*(*o*), and so on. In other words, for having color an object must have *a* color. Such coordinated – because exhaustive and mutually exclusive – properties $X_{P,i}$ may be called "manifestations" of *P*, or "determinates" of the "determinable" *P*. In this way the domain of *P* is partitioned into equivalence classes, where each $X_{P,i}$ identifies the class of objects *o* such that $has\_X_{P,i}(o)$. This process may be repeated by refining the partition: *has_color_red* can be considered a determinable in turn, whose determinates are *has_color_red_magenta*, *has_color_red_scarlet*, etc. Note that multiple criteria of partitioning are usually possible, and therefore that the set of determinates of a given determinable is not necessarily unique.

Whenever $X_{P,i}$ is considered as an element of X_P, the predicate $has\_X_{P,i}(o)$ is customarily written in the functional form of a *basic evaluation equation*

$P(o) = X_{P,i}$ in X_P

(for example, *color*(*o*) = *red* in {*red*, *blue*, ...}, thus highlighting the conditions of existence and uniqueness in X_P of the determinate $X_{P,i}$ of the determinable *P* for each object *o* in *D*(*P*). This equation formalizes the core information that can be obtained by a measurement: the object *o* has the property *P*, and in particular the property $X_{P,i}$ in the classification of *P* according to X_P (for example: *o* has *color*, and in particular has *color red* in the classification of *color* according to {*red*, *blue*, ...}). This justifies the usual understanding of *P* as a function that when applied to the argument *o* has the value $X_{P,i}$ in the set X_P, and provides a sufficient explanation of the somewhat confused related lexicon: both *P* and *P*(*o*) are considered to be properties (color is a property; the color of this object is a property that the object has), and $X_{P,i}$ in X_P is called a *value of a property*.

Measurement as representation

The idea of *measurement as representation* plausibly originates from here: as it was stated by Norman Campbell, who to measurement devoted a large part of his book *Physics: the elements* (1920), measurement is the process of assigning numbers to *represent* properties, where in fact "representation" and "assignment" are usual terms to designate functions. In this way the content of the basic evaluation equation is weakened, by omitting any condition about the *equality* of properties of objects and values of properties, and actually interpreting such equations as

$$representation(P(o)) = value$$

(or sometimes even as *representation*(o) = *value*, by thus absorbing the reference to the property in the representation). In this broad sense, measurement is considered to be the representation of properties of objects by means of values that are assigned according to some consistency conditions, the simplest of them being that distinguishable properties must be represented by distinguishable values, so that no information is lost in the representation. A basic set of such conditions was proposed by Stanley Stevens in his seminal paper *On the theory of scales of measurement* (1946), in which a sequence of scale types is identified, each of them characterized by an algebraic structure and a related condition of invariance. For example, if the properties x_i to be represented are not only distinguishable from each other but also ordered, then the values $f(x_i)$ by which the properties are represented must be ordered accordingly:

$$\text{if } x_1 < x_2 \text{ then } f(x_1) < f(x_2)$$

As a consequence, any further monotonic function g applied to the values preserves the information conveyed in the representation, i.e.,

$$\text{if } x_1 < x_2 \text{ and } f(x_1) < f(x_2) \text{ then also } g(f(x_1)) < g(f(x_2))$$

so that the property x can be represented not only by $f(x)$, but also by $g(f(x))$. For this reason g is called an *admissible scale transformation*.

Many physical properties are more than ordinal only, and admit a comparison by ratio (e.g., this is twice as much as that), sometimes based on an additive structure, such that objects can be composed according to their property and the composition has the features of a sum: $x_1+x_2 = x_2+x_1$, $(x_1+x_2)+x_3 = x_1+(x_2+x_3)$, and so on. At least in classical physics this applies to lengths, masses, durations, intensity of electric current, etc. In these cases a property may be singled out as the unit, mapped to the value 1 of the scale, so that all comparable properties are represented by multiples or submultiples of the unit, and the generic value $X_{P,i}$ in X_P can be meant as k u, i.e., the k-th multiple of the unit u (e.g., 1.234 m). In such a ratio scale any similarity,

$$g(f(x)) = k\, f(x), \text{ with } k > 0$$

is an admissible transformation, so that a scale factor k is sufficient to transform for example values in metres into values in inches.

The following table, adapted from Stevens' paper summarizes the main scale types and their defining conditions (where x is a property, v is a value, and g is a scale transformation).

<table 1 here>

This is the weakest, and therefore most encompassing, notion of measurement, also applicable to situations that someone would hardly accept as measurements at all (e.g., a measurement was performed because in one's opinion this is less than that and I labeled this as less than that). After Stevens this position was given a more rigorous mathematical formalization by the so-called *representational theories of measurement*, in which the fact that a structure of properties can be mapped into a structure of values, through a "representation theorem", is considered sufficient to make it a theory of measurement: if it is a consistent representation – in this algebraic sense of consistency – then it is a measurement.

Measurability

If the long scientific, technological, philosophical, and societal development of measurement is read in light of the representational approach, a clash emerges: while consistency in representation may be acknowledged as a necessary condition of measurement, its *sufficiency* would be questioned, by noting that, even if all measurements are consistent representations, not all consistent representations have been historically accepted to be measurements. Further possible constraints relate to the structure – whether quantitative or not only – and to the domain – whether physical or not only – of measurable properties. In both cases the root is plausibly to be found in the mathematics of classical Greece.

Can a non-quantitative property be measurable?

Ratios were so fundamental in Greek mathematics that "rational", i.e., "based on ratio", was meant also as "reasonable". And ratios were construed in terms of *measures*. Indeed, according to Euclid, in book v of *Elements*, "a magnitude is a part of a(nother) magnitude, the less of the greater, when it measures the greater", where Aristotle, in book v of *Metaphysics*, explained that "a quantum is a plurality if it is numerable, a magnitude if it is measurable. 'Plurality' means that which is divisible potentially into non-continuous parts, 'magnitude' that which is divisible into continuous parts.". On this matter distinguishing discrete vs continuous properties sounds today immaterial (hardly would someone claim that, say, the discovery that energy is quantized implies that it is countable but not measurable). Rather, important here is the nature of the entities whose measurability is considered. And that these concepts of magnitude and measure relate to mathematical, and not empirical, entities is clear from the fact Euclid himself considered that "a number is part of a(nother) number, the lesser of the greater, when it measures the greater". Two millennia later, Augustus De Morgan was keen in pointing it out: "the term 'measure' is used [by Euclid] conversely to 'multiple'; hence [if] A and B have a common measure [they] are said to be commensurable".

In the past the question whether the geometry of *Elements* is mathematical or empirical was possibly considered ill-posed, but with the break produced by the discovery of non-Euclidean geometries it became critical. And, despite some still existing stereotypes, the conclusion is clear: Euclid was concerned about the mathematical concept of *measure*, not the empirical concept of *measurement* discussed here. Hence, the source of the "special efficacy" of measurement cannot be the Euclidean concept of measure.

Can a non-physical property be measurable?

The emphasis on empirical entities of the experimental method promoted by Galileo Galilei led to reinterpret numerable pluralities and measurable magnitudes as physical quantities. In fact, a trace of the Euclidean focus on geometry remained in the traditional term “weights and measures”, as if weights (and pressures, temperatures, etc) were not measurable because only dimensional quantities are. For example, in his *Mathematical and Philosophical Dictionary*, 1795, Charles Hutton called “observation” the process of evaluating a temperature by means of thermometer: it was not a measurement, because only continuous geometric quantities were considered to be measurable. With such a strict implied meaning, it is not surprising that Otto Holder’s seminal paper (1901) *Die Axiome der Quantitat und die Lehre vom Mass* (translated as *The Axioms of Quantity and the Theory of Measurement*) maintained continuity and additivity as characterizing features of measurable properties. That most, if not all, psychosocial properties are not continuous, nor in fact additive, can be taken for granted: in no way the intelligence of two individuals is the sum of their intelligences, and so on (peculiarly, in those years physicists were discovering the discontinuity also of some physical quantities, at the fundamental level of quantum mechanics).

The measurability of such properties was the conundrum that led a committee appointed in 1930 by the British Association for the Advancement of Science to study “quantitative estimates of sensory events”. The conclusion was, substantially, that for a property to be measurable it must exhibit an additive structure or must depend functionally on one or more additive quantities (what Campbell had called *fundamental magnitudes* and *derived magnitudes* respectively). These conditions are possibly fulfilled by some psychophysical properties – along the line of the work started in the mid of XIX century by Gustav Theodor Fechner about the relationships between physical stimuli and the associated human responses – and in the mid of XX century triggered the development of methods to identify a hidden (interval) ratio structure in properties, particularly Rasch methods and the so-called “conjoint measurement” methods. But it also created a barrier that appeared impossible to overcome for many psychosocial properties, particularly in the context that we now call psychometrics.

A more encompassing concept of measurability, then?

The above-mentioned representational approach proposed a constructive strategy to overcome these limits, at the price of neglecting (i.e., abstracting from) the way the process of measuring is structured and actually performed, a choice that makes the approach general but also generic. For sure, quantitative evaluations convey more information than non-quantitative ones, as the theory of scale types shows; but the multiplicity of the scale types also highlights the conventionality of the position according to which only quantitative properties, and possibly only properties evaluated on ratio scales, are measurable. Even in the context of fundamental physical metrology additivity and continuity are not taken as necessary conditions of measurability any more, as witnessed for example by the *International Vocabulary of Metrology*, that assumes that also ordinal properties are (quantities and are) measurable.

The mentioned clash between the strict conditions implied by the historical conception of measurement and the loose perspective of representationalism is not solved: both demarcation – to distinguish measurement from other information-related processes – and qualification – to distinguish good from bad measurements – still require a synthesis.

Toward a theory of the process of measuring

Measurement is a process: hence, an actual theory of measurement, and not of measure, must be a theory of the process of measuring, and not only of the input-output relations that the process establishes. In reference to the metaphor of the black box, a theory of measurement has to deal also with what there is inside the box, and not only with what enters and leaves the box.

As a generic basic evaluation equation, $P(o) = X_{P,i}$ in X_P, shows (a quantitative example being *length*(*rod a*) = 1.234 m), measurement provides information about the property *P* of an object *o*, by stating that $P(o)$ is a given $X_{P,i}$ and not any other $X_{P,j}$, $j \neq i$, in X_P. There are fundamentally two strategies to produce this kind of information. First, the value of $P(o)$ can be computed via a function of already available values (for example, according to Newtonian mechanics the value of the length of the path traveled by a body moving at constant speed can be computed by multiplying the values of the speed and of the duration of the motion). Second, the value of $P(o)$ can be obtained by comparing the object *o* with other objects s_j, called *measurement standards*, each of them having a property $P(s_j) = X_{P,j}$ in X_P, so that $P(o) = X_{P,i}$ in X_P for a given *i* if $P(o) = P(s_i)$ (for example, *length*(*rod a*) = 1.234 m if there is a measurement standard s_i whose length is 1.234 m and is the same as the length of *rod a*).

Along the Galilean tradition, the "special efficacy" of measurement is a matter of somehow identifying the *measurand*, i.e., the property intended to be measured, and finding its value in the given scale in the here-and-now conditions of the process. To this goal some calculations may be useful or even required – for example to correct known environmental effects without an empirical intervention – but the first strategy is not sufficient: a pen-and-paper only evaluation is not a measurement, because a measurement requires the property-related direct or indirect empirical comparison of the object under measurement and the measurement standards that materialize the chosen scale (in the example above, in order to consider that the length of the path is measured – in what it is traditionally called an *indirect measurement* – and not just computed, the information about the speed and duration of the motion must be acquired in the actual conditions, something that a computation cannot provide).

The two key features to qualify a measurement derive from this. First, measurement results are expected to be *object related*, by reporting the outcome of the comparison of the object under measurement and some measurement standards. This explains the importance attributed to constructing and characterizing the behavior of measuring instruments, that have the goal of guaranteeing a sufficient object relatedness by removing the effects of all influence properties. Second, in most real-life cases measurements are expected to produce information that is socially transferable, and therefore interpretable in the same way by different individuals in different times and places, a condition that makes it *subject independent*. This explains the importance attributed to disseminating measurement standards and calibrating measuring instruments against them, so that the measurement results are metrologically traceable to the chosen scale, and in particular to the definition of the unit in the case of ratio quantities.

The quality of measurements and their results is fundamentally assessed about their object relatedness and subject independence, “objectivity” and “intersubjectivity” for short, through the several (usually quantitative) properties that characterize the process in its multiple facets: accuracy, specified as precision and trueness, in turn specified as sensitivity, selectivity, stability, resolution, and so on. The overall summary – traditionally related to measurement error – has been focused on *measurement uncertainty* in the last decades, hence with an encompassing epistemic emphasis on the information produced by the process and its trustworthiness: measurement is thus conceived of as an empirical and informational process that is designed on purpose, whose input is an empirical property of an object and that produces values of that property that are explicitly justifiable in their (good or bad) objectivity and intersubjectivity.

See also Accuracy; Applicability of Mathematics in Science; Data and Phenomena; Epistemology; Information Theory; Modeling; Scientific Realism; Statistics

Luca Mari

FURTHER READINGS

Campbell, N. R. (1920). *Physics: the elements*. Cambridge: Cambridge University Press.

JCGM (2008). *Evaluation of measurement data – Guide to the expression of uncertainty in measurement* (GUM). Sèvres: Joint Committee for Guides in Metrology, .

JCGM (2012). *International Vocabulary of Metrology – Basic and general concepts and associated terms* (VIM) 3rd ed (2008 version with minor corrections). Sèvres: Joint Committee for Guides in Metrology, www.bipm.org/documents/20126/2071204/JCGM_200_2012.pdf.

Kuhn, T. S. (1961). The function of measurement in modern physical science. *Isis*, 52(2), 161-193.

Kula, W. (1986). *Measures and men*. Princeton: Princeton University Press.

Mari, L., Wilson, M., Maul, A. (2021). *Measurement across the sciences – Developing a shared concept system for measurement*. Dordrecht: Springer.

Michell, J. (1999). *Measurement in psychology: A critical history of a methodological concept*. Cambridge: Cambridge University Press.

Rossi, G.B. (2014). *Measurement and probability – A probabilistic theory of measurement with applications*. Dordrecht: Springer.

Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, 103(2684), 677-680.

Tal, E., Measurement in Science. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*, plato.stanford.edu/archives/fall2020/entries/measurement-science.