34

MEASUREMENT

Hasok Chang and Nancy Cartwright

Introduction

Measurement is one of the most distinctive and pervasive features of modern science, but it is not easy to say what measurement actually is. Philosophers commonly define measurement as the correct assignment of numbers to physical variables. There are many difficult philosophical and practical questions about whether a measurement is made correctly and how we can know that it is. Various philosophical views surrounding these questions are discussed next; in the final two sections, we highlight concrete questions concerning the practice of measurement in the physical and the social sciences.

Epistemic questions

To the practitioner, the all-important question is whether measurements are carried out correctly. To the philosopher of science, that question acquires special significance in the context of the realism debate: does a measurement operation really measure what it purports to measure? Take one of the more controversial examples: does the IQ test really measure intelligence? To answer the question we need to consider not only whether the test results are in line with what we intuitively understand as "intelligence," but also whether the presumed quantity really exists. Two broad positions can be identified about the nature of measurement: one treats measurement methods as definitive of the concept; the other takes measurements as methods of finding out about objective quantities that we can identify independently of measurement. These positions could be characterized, respectively, as *nominalism* and *realism* about measurement.

The core of nominalism is a rejection of the realist question about the correctness of measurement. Within nominalism, we can again distinguish two positions. The more extreme is *operationalism*, which maintains that the meaning of a concept is fully specified by its method of measurement, implying that each measurement operation defines its own concept; consequently, it becomes a tautology that any measurement operation is the correct one for the concept associated with it. Operationalism is commonly associated with the American physicist Percy W. Bridgman, who once declared: "In general, we mean by any concept nothing more than a set of operations; the concept is synonymous with the corresponding set of operations" (1927: 5). Bridgman later regretted having formulated such a narrow view, distancing himself from the term "operationalism" or "operationism." Instead, he emphasized another strand that was always present in his writings: the usefulness of analyzing scientific practices and epistemic situations in terms of operations. Among other benefits, such operational analysis can reveal divergences in practice that careless linguistic and mathematical habits conceal. For example, consider the diversity of operations underlying the notion of "length": in everyday circumstances, we have operations like lining up meter-sticks against solid objects; measuring atomic dimensions requires putting together some complicated equations of electromagnetic theory or quantum physics with some observable quantities; measuring astronomical distances necessitates a host of different operations depending on the scale, starting with the measurement of the time light takes in reaching an object and traveling back after being reflected. According to operationalism, there are as many concepts of length as there are different types of operation used for measuring it.

The less extreme nominalist view is *conventionalism*, according to which we are free to choose by agreement the *correct* measurement method for a concept. Here it is useful to make a distinction between *definition* and *meaning*. We do not have to be close followers of the late Wittgenstein to admit that the meaning of a concept derives from all the different ways in which it is used. When we fix on a definition of a concept, the intention is to regulate its uses; the definition allows us to judge whether the use in question is correct or not. Pure operationalism defines concepts in terms of measurement operations, and then reduces down their meaning to such operational definitions. Conventionalism does not conflate meaning and definition but allows a convention, for example an agreed measurement operation, to regulate the use of the concept. Because nature does not dictate the correct method of measurement, we are left with convention as the highest epistemic authority. A prime example of conventionalism is Henri Poincaré's discussion of time measurement (2001 [1913]: 215): "time should be so defined that the equations of mechanics may be as simple as possible. In other words, there is not one way of measuring time more true than another; that which is generally adopted is only more *convenient*."

Nominalist positions are motivated partly by the recognition that many of the entities, properties, and relations that interest scientists are unobservable. This is not only about physics and chemistry venturing into the microscopic realm. One of the influences that pushed Bridgman toward operationalism was Albert Einstein's exposé of the impossibility of determining absolute simultaneity at a distance (Bridgman 1927: 1–9). It was an immense shock to many physicists and philosophers to realize that they had taken for granted the meaningfulness of the Newtonian notion of *distant simultaneity*, whereas critical thought should have made it obvious that it cannot be determined without adopting one measurement procedure or another, either of which lacks absolute justification. Bridgman, with his operational analysis, sought to "render unnecessary the services of the unborn Einsteins" (*ibid*.: 24).

Realism denies that measurement methods are definitive of concepts. For the realist, measurement is an activity aimed at discovering the true value of a specified quantity that exists independently of how we measure it, and the question of the correctness of method is certainly not vacuous. "Does operation O measure concept C correctly?" is a question that must be taken seriously – and answered in the affirmative – by any empiricist who wishes to test the truth of any theories that involve C.

Within the sciences and even in philosophy, there is widespread *naive realism* about measurement that consists in the assumption that our familiar measurement methods correspond correctly to the concepts specified by our theories. In many cases, the situation is far more complex. For example, does the standard mercury thermometer measure temperature correctly? In common conception (though not in modern expert practice), the mercury thermometer is a mercury-filled cylinder of uniform bore, calibrated at the freezing- and boiling-points of water to read 0°C and 100°C, with the scale between those fixed points divided up uniformly and extrapolated beyond them. Such an instrument would give correct temperatures only if the mercury expands uniformly with temperature. How can we test that assumption? We need to monitor how the volume of mercury varies with real temperature; if the volume is a linear function of temperature, then our mercury thermometer is correct. But how can we get the real temperature values without already having a thermometer that we know we can trust, which is just what we are trying to obtain?

This problem of justification is common to all measurement methods based on empirical laws. Hasok Chang (2004: 59) has dubbed it "the problem of nomic measurement." We seek to determine quantity x via another, more easily observed, quantity, y, with the help of an empirical law expressing the former as a function of the latter: $x = f(y)$. In order to test our expression for f empirically we would need to observe values of x and y, but without f already established we cannot determine the x-values empirically. There are two obvious ways of trying to avoid this problem. First, determine the x-values by another measurement method; this only postpones the problem, as we would need to ask how that other method is justified. Second, derive f from a more general theory; this is not straightforward, either, as we would need to know that the theory was empirically justified, which would inevitably involve measurements of other unobservable quantities, if not x itself.

The problem of nomic measurement is a sharp manifestation of the more general problem of the theory-ladenness of observation. There are extreme types of theory-ladenness in modern measurements of many quantities, for example, very low temperatures, properties of elementary particles, and distances to faraway astronomical objects. Pierre Duhem (1962 [1906]) long ago noted how the necessity of justifying the workings of measuring instruments leads to holism in epistemology. In order to defend realism about measurement, one needs to have a way of dealing with theory-ladenness and holism in general. A mild version of operationalism can be seen as an attempt to avoid holism by avoiding theory-ladenness. If empirical concepts can be defined by well-specified measurement operations, observational data can be fixed without reference to theories and be made secure, while theoretical concepts and laws fluctuate and develop. Whether there are theory-free operations that can support sufficiently useful empirical concepts depends on the circumstance. Herbert Feigl (1970) noted that our most basic measurement operations are grounded in middle-level regularities that seem to have a remarkable degree of stability, such as Archimedes's law of the lever and Snell's law of refraction.

Whether nominalist or realist, those who practice measurement tend to be concerned about precision. In common parlance "precision" is often confused with "accuracy." *Accuracy* is a realist notion about whether measurement results agree with the true values; *precision* is a concept that is meaningful to the realist and nominalist alike, as it indicates merely how specific a measurement result is. One might say that precision is a necessary but insufficient condition for high accuracy. True precision requires consistency of results when repeated measurements of the same quantity are made. Different authors use different terms to express the accuracy–precision distinction. For example, statisticians commonly distinguish *validity* from *reliability*; the distinction also maps on to that between *error* and *uncertainty*. In some circumstances, the same operational measures or statistical data-processing techniques serve the goals of both accuracy and precision.

Some problems of measurement in the physical sciences

Quantification

Steeped in modern scientific thinking, we tend to think of all physical properties as numerical quantities amenable to measurement. It can be a shock to learn the list of physical concepts that used to be considered qualities to which numbers could not be attached. For example, Alistair Crombie (in Woolf 1961: 21–4) explains how fourteenth-century Oxford scholars struggled to quantify velocity, which had been considered by most Aristotelians as an unquantifiable quality. Another Aristotelian quality was heat, which was quantified during the seventeenth and the eighteenth centuries into the distinct modern concepts of *temperature* and the *quantity of heat*. Quantification of many other concepts in physics and chemistry followed. Acidity (and alkalinity) presents an interesting case: the modern measure of it is expressed in *pH* values, based on the concentration of hydrogen ions. That quantification of acidity made the meaning of the concept more specific than it had been and also ruled out certain previous concepts of acidity. A more extreme case of such narrowing and changing of meaning through quantification is that of color via wavelength.

Attempts at quantification do not always succeed, even in the physical sciences. One example is chemical affinity. Between the late eighteenth century and the early nineteenth century there were various schemes for measuring the strength of affinity between different chemical substances. This was an entirely sensible enterprise, since much chemistry in that period was based on ordinal rankings expressed in affinity tables which explained why certain combinations happen in preference to others. It was, therefore, a natural hope that coherent numerical values could be assigned to affinities. In this case quantification turned out to be a mirage, as further investigations revealed that even the ordinal rankings were not robust, being subject to flipping depending on external circumstances such as heat and wetness. Color is another interesting example. Psychologists studying color perception by mapping the perceived degrees of closeness between various hues found that the perceived relationships could be adequately represented only in a two-dimensional color circle, which cannot be mapped onto the linear spectrum of wavelengths.

The improvement of precision

Practically speaking, the best advertisement for quantification is precision. On the whole, the physical sciences have been extremely successful in improving the precision of measurements. Observational astronomy was probably the first field of science that developed specialized instruments and practices designed to increase precision, showing impressive achievements already by the sixteenth century, thanks to the likes of Tycho Brahe. By the mid- to late eighteenth century other physical quantities began to be measured with great precision. Fine balances for weight measurement were constructed, allowing Henry Cavendish, Antoine Lavoisier, and others to weigh gases and Count Rumford to argue that heat was not a substance because it had no detectable weight. Mechanical and pendulum clocks were developed well enough to show that the length of the day (from noon to noon) was not constant, and John Harrison with his famous marine chronometers led the pack of horologists searching for a method of making accurate longitude determinations at sea. Surveying techniques were sufficiently developed for teams of French scientists to determine the length of 1° of arc on different parts of the Earth with precision; this helped to settle the debate between Newtonians and Cartesians about the shape of the Earth, and also served as a basis for the definition of the meter adopted during the French Revolution. Charles Augustin Coulomb developed a torsion balance for the precise measurement of force, which he used in his investigations in electrostatics; Cavendish used a similar arrangement to measure the gravitational force between terrestrial objects. For the measurement of small lengths, micrometers were developed, and the engineering of other precision instruments depended crucially on the exact control of the dimensions of parts. Over the nineteenth century, a culture of precision took hold of experimental physics as a whole, to which the contributions of Victor Regnault were significant; gradually many other laboratory-based sciences followed suit.

Despite this impressive list of achievements, there is a deep epistemological question about how it is possible to increase precision, which can be illustrated with the case of temperature. If we only have thermometers that measure down to 1° to begin with, how will we be able to judge whether a new thermometer that measures down to 0.1° is correct? Relying on theory creates the same difficulties discussed in section 2. If the justification is empirical, then a lower-precision instrument is being asked to underwrite a higher-precision instrument. This is a general problem, to which there is no simple, realist solution. In the iterative development of precision, there is at each step a choice to be made between competing higher-precision standards, each compatible with the previously accepted lower-precision standard. How that choice can and should be made are serious philosophical and practical issues (Chang 2004: Chs 3 and 5).

The choice of convention

Once we allow a degree of nominalism about measurement, interesting issues emerge about the choice of convention. The competition between solar time and clock time gives a good illustration (Landes 1983: 122ff.). Clock time, which declares the movement of the Sun irregular, appeared absurd to those who regarded astronomical regularities as the most important and even definitive aspects of the meaning of time. As noted in section 2, a definition is an attempt to regulate the divergence of meaning. Any concept familiar to general society, such as time, is bound to have a multifaceted meaning. The measurement of such a concept with any precision is likely to sacrifice or alter aspects of the meaning. In the case of time, any quantification at all is a departure from some aspects of the inner experience of it, as Henri Bergson argued.

There have been many debates about the choice of measurement unit and scale – some of them quite heated – as between Fahrenheit and Celsius, or metric and imperial. Philosophers may smile at these tussles over what seems an arbitrary issue, but the force of custom is considerable, as shown by the failure of the decimal clock and the ten-day week proposed, along with the metric system, during the French Revolution. Moreover, a unit is often not just about the size of the quantity we take as the base of counting. The choice of unit and scale is often tied up with the choice of measurement method, which is in turn based on substantive assumptions. For example, measuring distance in light-years is based on the assumption that the speed of light is constant. Similarly, it is too simple to say that *degrees Kelvin* is just *degrees Celsius* minus 273.15°. Lord Kelvin's absolute temperature concept sprang from his desire to avoid reference to any particular material substance in the definition of temperature, and it was based on the abstract theory of thermodynamics for that reason. The traditional Celsius scale was based on the system of two fixed points and relied on the assumption of the linear expansion of mercury. (In fact, the original temperature scale of Anders Celsius was *upside-down*, with 0° denoting the boiling-point of water and 100° the freezing-point; it is interesting to speculate about what exactly Celsius was trying to measure on that scale.)

Some problems of measurement in the social sciences

As Max Weber taught us, the social sciences face a number of special problems with measurement that are more severe than those in the physical sciences. We discuss some of the more pressing issues here.

(1) Physical sciences look for exact laws involving unambiguously defined and measurable concepts, and they can adjust their choice of concept to serve this aim. If one candidate proves inconvenient, it can be replaced by another. Consider the acceleration of falling bodies, which go faster the longer and farther they fall. Medieval scholars tended to define acceleration as the increase of velocity as a function of distance traveled by the body. Modern physicists prefer to use dv/dt, the rate of increase of velocity with time; this formulation has many advantages, including its

role in Newton's second law of motion. The social sciences have no such latitude. They are supposed to help us understand the behavior of the factors we are interested in, which may not figure in strict laws nor be exactly measurable.

Measurement in the social sciences involves two kinds of activities: providing a theoretical definition for the quantity of concern, and devising and defending empirical procedures for determining when the concept applies in the world. The *theory of measurement* (see Suppes 1998 for an accessible introduction) concerns the first and, although its strictures apply equally in the natural and social sciences, social scientists are more attentive to its demands. The first task within measurement theory is to provide a mathematical representation of the targeted concept so that it can be integrated into a theory with an existing set of concepts. In the falling-body example above, both concepts of *acceleration* can be equally integrated with existing concepts.

The second task is to provide a *representation theorem* to show that this representation is adequate. A representation theorem first provides a set of characteristics taken to be true of the targeted concept, and then proves that the concept as defined has those characteristics. Consider economic freedom. We talk loosely of economic versus political freedom, of negative versus positive freedoms, and the like. Can economic freedom be defined more exactly in the framework of, say, social choice theory? The simplest idea is a pure cardinality measure that identifies the degree of economic freedom agents have with the number of options available to them. Is this a good definition? Suppose we agree that economic freedom has some basic features: for example, if one set contains every option that a second contains and more, the first offers more economic freedom that the second. In a good exemplar of measurement theory at work, Pattanaik and Xu (1990) provide axioms describing three such features, then prove that an ordering among sets of options satisfies those axioms just in case it orders the sets according to their size. Later writers provide more nuanced definitions. In each case measurement theory requires that the definitions be defended by a representation theorem.

Measurement theory regulates only half the job: once a concept has been defined within a theory, empirical procedures are required to tie it to the world. How, for instance, do we measure the size of someone's economic choice set? In psychometrics these two stages are often collapsed into one. Suppose a set of measurement procedures for a concept is on offer, say a questionnaire, to determine how depressed one is. Psychometrics offers a number of tests designed to provide evidence about whether the questionnaire is indeed a measure of depression. The analysis, defense, and improvement of such procedures are among the central tasks of methodology of the social sciences.

(2) Even if we assume that our social concepts pick out real quantities, there are other difficulties in the attempt to provide measurement procedures for them:

- Measurements of psychological states will always be indirect. Even honest and attentive self-reports cannot be taken as reliable without more corroboration.
- For the purpose of comparisons, measures and measurement procedures are required that can be applied across locations, populations, economies, and cultures. This

often results in measures that lose information – measures that are far from the best procedures that could be devised in the separate groups – and the more local measures often give dramatically different results from the more universal ones. Also, for theory-testing we need separate procedures that measure the same univocal concept, but for use we generally need a variety of purpose-specific concepts, each with measurement procedures appropriate to it. The two demands pull in opposite directions.

- Because people are self-conscious and reflective and because social institutions are often designed to be plastic and responsive to their environment, it is often difficult to design measurement procedures that do not significantly disturb the measured systems.
- Moral, political and cultural norms severely restrict the kinds of measurement operations that can be performed on people and their social institutions.
- We often want to measure aggregate and ambiguous concepts, like the total value of goods and services produced in a country. How do we do so since we cannot count them all; and how do we decide what is to be counted? For instance, is household labor to be included?

(3) Measures in social science are often not value-free despite our best efforts. Very frequently, they make sense as measures only in relation to certain values or purposes. This may be obvious in a case like the human development index, which includes life expectancy, level of education, and GDP. Should it include a measure of political freedom as well? That presumably depends on whether political freedom is accepted as a constituent of human flourishing.

The intrusion of values or purposes may be less expected elsewhere, but it seems exceedingly difficult to avoid. Consider the recent Boskin Commission proposals in the US for revising the consumer price index (CPI). One proposal argued that the price for many goods is overestimated because it is based on samples from retail stores, whereas the goods tend to be much cheaper in outlet stores, which are not properly represented when prices are sampled. But, as Julian Reiss (forthcoming) argues, adjusting the CPI in this way will disadvantage the elderly, those without cars, and other groups who have poor access to outlet stores, which are generally far from town centers.

A stock response to these problems urges that decisions involving value-laden choices in the construction of a measure be given to users of the measure – policy-makers of all sorts who will use the measure in their deliberations. This has major drawbacks. First it leads to a proliferation of measures which become difficult to understand and keep track of; we also get the same problems of theory-testing and comparison discussed already with respect to universal versus purpose-built measures. Second, it is an extremely difficult strategy to execute. Consider poverty measures. Perhaps a legislative body or the populace is willing and able to think about whether the measure should be absolute or relative, and, if relative, relative to what. Should we set the poverty line at two-thirds of the median income? Should we count households or individuals? How should we weight individuals in a household? Those decisions both affect different groups in different ways and also can dramatically change the assessment of how much poverty there is and the poverty-rankings among different regions. To understand the impact of those decisions requires much thought and more economic and social knowledge than even experts have, let alone those who want to use the information. Here again is a problem that makes designing measures in the social sciences far more difficult than in the natural sciences.

See also Evidence; Scientific method; Social sciences; Values in science.

References

Bridgman, Percy Williams (1927) *The Logic of Modern Physics*, New York: Macmillan.

Chang, Hasok (2004) *Inventing Temperature: Measurement and Scientific Progress*, New York: Oxford University Press.

Duhem, Pierre (1962 [1906]) *The Aim and Structure of Physical Theory*, New York: Atheneum.

Feigl, Herbert (1970) "The 'Orthodox' View of Theories: Remarks in Defense as well as Critique," in Michael Radner and Stephen Winokur (eds) *Analyses of Theories and Methods of Physics and Psychology*, Minneapolis: University of Minnesota Press, pp. 3–16.

Landes, David (1983) *Revolution in Time: Clocks and the Making of the Modern World*, Cambridge, MA: Harvard University Press.

Pattanaik, Prasanta and Xu, Yongsheng (1990) "On Ranking Opportunity Sets in Terms of Freedom of Choice," *Louvain Economic Review* 56: 383–90.

Poincaré, Henri (2001 [1913]) "The Measure of Time," in *The Value of Science: Essential Writings of Henri Poincaré*, New York: Modern Library, pp. 210–22.

Reiss, Julian (forthcoming) *Error in Economics*, London: Routledge.

Suppes, Patrick (1998) "Measurement, Theory of," in E. Craig (ed.) *The Routledge Encyclopedia of Philosophy*, London: Routledge; available: http://www.rep.routledge.com/article/Q066?ssid=388873355&n=8.

Woolf, Harry (ed.) (1961) *Quantification: A History of the Meaning of Measurement in the Natural and Social Sciences*, Indianapolis, IN: Bobbs-Merrill.

Further reading

Useful philosophical discussions about measurement in various sciences can be found in John Forge (ed.) *Measurement, Realism and Objectivity: Essays on Measurement in the Social and Physical Sciences* (Dordrecht: Reidel, 1987). Woolf (1961) gives very interesting historical views on quantification in the natural and the social sciences. Broader historical and cultural perspectives on precision measurement are provided in M. Norton Wise (ed.) *The Values of Precision* (Princeton, NJ: Princeton University Press, 1995). For those interested in following up on issues concerning economic measurements, an excellent place to start is Judy L. Klein and Mary S. Morgan (eds) *The Age of Economic Measurement* (Durham, NC, and London: Duke University Press, 2001). Broad surveys of measurements in a wide variety of fields can be found in David J. Hand, *Measurement Theory and Practice: The World Through Quantification* (London: Arnold, 2004), and Herbert Arthur Klein, *The Science of Measurement: A Historical Survey* (New York: Dover, 1974). Those interested in studying formal theories of measurement, introduced in Suppes (1998), can refer to David H. Krantz et al., *Foundations of Measurement*, 3 vols (New York: Academic Press, 1971–90).