Problems in theory of measurement today

L. Gonella

Istituto di Fisica Sperimentale, Politecnico di Torino, C. Duca degli Abruzzi 24, 10129 Torino, Italy

This is a reissue of a paper which appeared in ACTA IMEKO 1979, Proceedings of the 8th IMEKO Congress of the International Measurement Confederation, "Measurement for progress in science and technology", 21-27.5.1979, Moskow, vol. 1, pp. 103–110.
The main questions to be answered in the theory of measurement are presented, and their formulation analysed in the classical approach and in a new operative one: the scope and aims of the theory itself, the concepts of measurable quantity, measurement results, error, uncertainty, the role of time. The paper is meant to define problems as the first step toward their solution.

1. SUBJECT AND SCOPE OF THE THEORY OF MEASUREMENT

To ask what is (or should be) the subject matter of the theory of measurement is not an idle question, since what seems to be the first problem today is to realize that there are indeed basic problems to solve and to identify then. Too many people feel with undue complacency that the theory of measurement is a perfected structure, needing only peripheral work. Others feel, instead, that fundamental questions lie unanswered, or even unasked, being pre-empted by undiscussed assumptions; framing such questions, that I think are to be asked and answered from the operative standpoint [1–4], is the aim of this paper.

The purpose of a theory is to describe reality in a coherent framework – in our case to provide a coherent description of the valid praxis of measurement, suited to its whole range. It is apparent that the current theoretical framework is not up to this task: the rank confusion prevailing in the field is widely recognized and lamented. Just try and compare the definitions given in the current literature of basic concepts as value, error, measurement, measurable quantity, etc. (I even saw the lecturers on Metrology of an International School declining to hold a seminar on definitions as meaningless in today's mixed-up situation [3]). There is a widening gap between the praxis of measurement and its theory: our measurements agree with one another much better than we agree in talking about them.

Curiously enough, this is not felt as a flaw in the theory: for most people it is just a matter of 'terminology' and the problem is shunted to nomenclature committees, which face up an impossible task. They are supposed to choose the 'right word' for designating well-understood, undisputable concepts, but these ones turn out to be hazy or unfitting; a formal redefinition is called for, and in so doing unwitting theoretic work is done: what else is the core of a theory but a cross-connected list of definitions?

The 'current' theory of measurement is embodied in the definitions, remarks and assumptions (both discussed and taken for granted) of textbooks on one hand and normative literature (standards, codes of practice, etc.) on the other. There is no parallel of the latter source in the other fields of science, unencumbered by the strong normative implications of the measurement problem. What happens is that the 'normative' authors feel that it is not up to them to discuss the theory, but only to codify the approved practice using the language of the main line of the approved textbooks; the latter, on their hand, cannot teach differently from the main line of the normative literature without rendering a bad service to their readers. The vicious circle thus generated makes it difficult to update the theory; its haphazard growth disguised as lexical adjustment or conventional ruling cannot touch the basic tenets, while the due concern over its responsibility may well transform some normative body into a low-pass filter against new ideas. This is one reason, I think, why the current theory of measurement is still based on the framework masterly set up by Helmholtz [5] on basic ideas of Euclidean lineage, which makes it the only field of science treated as yet in full 19th—century terms. A century passed, which shook the very foundations of mechanics and geometry, but new basic concepts like those of signal and

information, of quantum indeterminacy, etc., do not seem to have found their proper place in the frame of the measurement theory. Meanwhile a more and more sophisticated instrumentation calls for a refined treatment in modern setting; we need a comprehensive theory able to range from the carpenter's rule to the microprocessor, and the latter can hardly be dealt with as an extrapolation of the former.

As technical-scientific people are shy of meddling with 'philosophical' issues concerning the meaning of what they do, and grew accustomed to shift most basic issues into terminological channels, the bulk of their work is devoted to mathematical techniques for data elaboration, mostly concerning the probabilistic error analysis. Somebody is therefore inclined to identify the theory of measurement with this analysis (blanket statements as "any physical quantity is just a statistical variable" are often heard among physicists), which represents a drastic reduction in scope for the former and leaves the latter without bases (such position is also due to the general, very disputable, inclination to identify 'theory' with mathematics). The debate on the basic concepts of measurement has been left mainly to scholars in non-physical sciences [6–9]: almost everybody seems convinced that only there some problem is left, while in physics and the related technical fields all is nicely settled on Helmholtz's lines [8]. In the minds of many people Stevens's ideas are connected with non-technical measurements, though his stand against Campbell [6] on psychophysical measurements could well have been taken by any technical scholar on several engineering measurements (e.g. hardness). Marvel was expressed for the neglect in which the field is left [7], and it was noted that fundamental issues about measurement in physics are unsolved or plainly ignored [9].

Measurement is a knowledge-gathering technique. What its theory is to be concerned with depends on what we expect from a theory. From the operative standpoint the measurement theory is the total body of definitions and rules concerning measuring activities, formulated with the aim of providing a guide to get consistent results. From another it is the argumentation that justifies the measurement itself, thus sanctioning what can be measured and what cannot. The issue of the scope of the measurement theory is coextensive with the question of what is a measurable quantity.

2. MEASURABLE QUANTITIES

In an operative approach we may say a measurable quantity is anything we can describe in an objective and consistent way; the problem then is to define 'objective and consistent' in order that we can measure things to our satisfaction and discriminate against illusory knowledge. The classical approach starts from the other end: the consistency is looked for in the formal properties of the class of mathematical entities chosen for the description, the measuring procedure is modelled on a formal operation of this class, and a measurable quantity is what fits the model; the problem then is to justify the measurement of the other things that in practice we measure but do not fit directly in the pre-established model.

The universal (though not necessary [9]) choice is to map the measurable quantities on the class of the real numbers, setting up a homomorphism between the empirical juxtaposition of the measured objects and the addition of the numbers assigned to measure them. The obvious paradigm is the measurement of length. All quantities endowed with the 'additive' property of the length, called fundamental or extensive (terminology and connotations depend on the author), may be measured directly; the others must be somehow 'derived' from them; if some kind of 'derivation' is not devised the quantity is not deemed and is excluded from the scientific realm. This way one gets a very good logical consistency for the fundamental quantities, but at the expense of generality, and the approach itself breeds the inclination to reduce the scope of the theory. When a theory meets phenomena which do not fit, either one changes the theory or puts an 'off limits' sign on the phenomena; the more clean-cut is the theory, the stronger the temptation to shrug off what does not fit. The 'derivation' criteria, either of the quantity or of the measuring procedure, are debatable. Considerable ingenuity was spent for deriving the measurement of temperature; other quantities (e.g. hardness) were not so lucky, perhaps because they were not so firmly embedded in the core of physics, and are therefore in a limbo – someone calls them "pseudo-quantities", which does not deter people from measuring them, but only hampers coherent standards. Likewise, nuclear physicists keep using counters as measuring instruments while others show that counting is not a measurement [7]. Something is wrong: either the people who keep on measuring "pseudo-quantities", or the arbitrary requisite of an 'additive' property that so many measurands simply do not possess. One must remember that once people believed that all measurements ought to be referred to mass, length, and time: much of this kind of philosophy is still attached to the measurement theory. The 'derivation' problem is circumvented by the idea of different 'scales of measurement', with different relational properties, but many people do not accept all 'scales' as 'measurement', and others distinguish between a strict sense and a wide sense of measurement. It is also disturbing that the vectorial character of physical quantities cannot be accounted for in these frameworks.

Anyhow, if 'measurement' is a procedure of limited applicability, other consistency rules, beside another label, must be given to the knowledge-gathering procedure applied to the "pseudo-quantities", and an overall treatment is in order. Is a distinction between strict and wide sense of measurement meaningful, or is it not a question of different types of measurable quantities? The idea itself of looking into mathematics for 'justifying' a measuring procedure, for finding out whether or not we are allowed to measure a thing, is disputable: is the theory of measurement a branch of mathematics (albeit applied), or is it a branch of the natural and technical sciences (using suitable mathematical tools)? There is a difference, which lies in where we look for the ultimate criteria of consistency.

3. THE DATA SET REPRESENTING A MEASURAND

Granted that measurement is, in Finkelstein's well-chosen words, "the process of assignment of numbers to attributes ... in such a way as to describe them" [10], we have yet to decide in which way numbers are to be assigned. In the classical theory the question is not even posed; to represent the measurand one real number is assigned, tied to a unit of measure: the value of the measurand in the chosen scale of measurement. The geometric paradigm indeed leaves no doubt that "measurement demands some one-to-one relation between the numbers and the magnitudes in question" [11]. The fact is that with a single real number to represent the measurand no consistency can be obtained in different measurements of the same measurand, owing to what Eisenhart aptly calls the "cussedness of

measurement" [12]. The difficulty is bypassed by assuming that the 'true' value – the one number that would make all measurements fit – exists, but is unknown and unknowable; individual measurements yield values differing from it by an 'error' due to instrumental 'imperfections'; an analysis of these measurement errors is required to assess the interval within which the true value is supposed to lie. This way theory of measurement and theory of errors are separated: the former is free to pursue its way unencumbered by questions of uncertainty, left in care of the latter. Somebody finds it a good arrangement, claiming that the principles of measurement cannot depend on 'technical' issues of uncertainty [8]; others feel that 'essential difficulties in the logic of measurement arise from uncertainty' [10]. To postulate the solution is no way to solve a problem [11]: this course was taken only on the strength of the geometric paradigm.

We should ask: is the real number the proper mathematical entity for representing measurands? The real number was invented to solve problems in pure mathematics born of the Euclidean mensuration problem, i.e. expressing with a number the proportion of two segments of any geometric figure; the apodictic certainty offered by geometry is so attractive that one forgets it is due to the abstract definition of the geometric figures, that do not belong in the real world. But measurement is meant to describe the real world, and if it follows too closely the tracks of geometry it risks to become itself involved in idealized measurands. The real number is by definition the limit of a convergent sequence, but in measurement there is no such a thing as a sequence with a definite generating rule. We may also ask: is the mathematics of continuum, of which the real number is the main pillar, the proper tool for representing the physical world? Atomic and nuclear physics did show matter to be granular; quantum mechanics added a granularity of its own and also an indeterminacy deeply embedded in the very fabric of things physical. This is not the setting meant for real numbers and the mathematics of continua. It would not do to have a different theory of measurement for the macroscopic and microscopic worlds. Besides, everybody knows that any real macroscopic measurand is endowed with an intrinsic uncertainty inherent to its very definition ("you don't measure bricks with a micrometer"); with more and more precise measurements of a given measurand one does not get closer to a real number expressing its value, as the classical theory assumes mimicking sequences, but rather at a certain point one finds oneself measuring something else, because the measurand's definition has exploded into a finer structure. The implications of this commonplace experience are lost if one postulates descriptions by real numbers, and one also wonders whether possible ties between the quantum indeterminacy of microphysics and this 'definition indeterminacy' of macrophysics were not obscured by the slant toward geometric sharpness taken by the theory of measurement.

If we look at things from the operative standpoint, we cannot help noting that measurements actually yield numerical intervals, and that upon such intervals we must reason to judge the consistency. Even a 'perfect' instrument could not overcome the intrinsic uncertainty of a real measurand. The mathematical entity we are actually working with is a set of numbers, and there is nothing wrong in representing a measurand with a full set of numbers tied to a unit of measure: it is yet "assignment of numbers ... in such a way as to describe it". This set assigned to describe the measurand (let us call it a "value-span") would include in the description the uncertainty of the measurement, thus reuniting theory of measurement and 'theory of error'. The one fundamental requirement for consistency is that different valuations of the same parameter yield 'the same' result: it is matter of defining rigorously what we mean by 'the same' and follow suit. Two measurements are consistent if the assigned value-spans overlap: this translates in formal terms the practical judgment of consistency "to stay within the error". The algebra of sets is to substitute the algebra of real numbers in operating on measurement data, starting with the replacement of the basic relation 'equal' with 'not disjoint' [1–4]. A new Italian Reference Standard on basic measurement terminology [13] follows this line, that was found easy to practice.

One may ask which special properties should have the sets assigned to describe measurands. In particular, the boundary of these sets cannot be sharply defined, if the pitfalls of the real-number concept are to be avoided. A suitable specialization of the 'fuzzy set' concept might well turn out to be the proper mathematical tool for the job [1–3, 14].

4. ERROR, UNCERTAINTY, INDIFFERENCE

The problem posed by the cussedness of measurement is tackled by the classical theory in terms of 'error'. This concept rests on the idea that the instrument ought to indicate directly the true value of the measurand but indicates instead a wrong one; the underlying assumptions (derived straightly from the model of measuring length with a graduated rule) are: (a) the measurand is described by a single real number; (b) the instrumental indication is a value assigned to the measurand. On this basis the questions to ask are: why is the instrument wrong? how much is it wrong? A full answer is impossible: we cannot hope to know the error a priori, we can only judge a posteriori that it is smaller than a certain amount. This 'maximum error' is thought to be determinable by combining the different errors due to various causes, separately analysed, and an a priori distinction is made between 'random' and 'systematic' error; the former may be treated exactly by statistics, but the latter turns out to be an elusive concept [12] that calls for an educated guess. The distinction has some heuristic value, beside historic reasons, but adds more problems than it solves: without a sum rule it is no help for judging the consistency of different measurements, and no logically sound sum rule can be given for such essentially heterogeneous components (as shown by the fact that at least four competing rules are now used). Such care was paid in distinguishing between 'precision' and 'accuracy' that some technical language even lacks a word describing their cumulative effect; so it happens that after defining different errors one often drifts to speak of 'uncertainty', without however defining it.

The problem may be tackled from a different standpoint if we drop the basic assumptions of above, remembering that the theory had been cast in that mold long before the 20th-century concept of signal were developed. If we look at the instrumental indication as at a signal bearing information on the measurand, and allow that we are to assign a data set describing the latter (not necessarily a single number) by connecting this information with other information pertaining to the instrumental characteristics and to the 'influence quantities' affecting the measurement, then the questions to ask are: how do we describe the instrument output? which relation holds between this output and a value-span suited to describe the measurand? This way we keep much closer to the actual practice, and may account in the same framework for all

instruments (allowing for the evolution of the output signals from the reading of graduated rules to the string of digits out of on-line computers). This approach calls for a clear distinction between the terms referring to the description and those referring to the instrument's output. Much confusion is thus avoided that the classical approach brings about: many technical languages, e.g., use the same term for both the operation of measurement and the resulting description of the measurand, following a mathematical usage justified only by the uniqueness of mathematical solutions; the statistical treatment of 'errors' is also often confused with that of random quantities [4]. The attention is brought on the calibration procedures and the interplay of instrument, measurand, and environment.

It is now possible a question that the classical approach cannot ask, i.e., is there a relation between the uncertainty of the measurement and the information gained on the measurand? The error concept reflects only the inability of the instrument to supply the information, assumed to be contained in the true value: the measurement is treated in an all-or-nothing way, and we cannot tie a larger error with less information. The uncertainty concept has uncertain meaning. Taken as 'the maximum allowed error' of the instrument, a given uncertainty associated to the value indicated by the instrument means that we are uncertain on where the true value is in the interval thus defined, and the width of the latter might be construed qualitatively as information on the measurand. On the other hand, if we treat this interval as a set assigned in its whole to describe the measurand, this means we are indifferent on which element of the set represents the measurand: its description is the whole set and two measurands cannot be distinguished from each other if their value-spans overlap; then an instrument which in correspondence to its output signal is able (through its calibration operator) to assign a narrower value-span is also able to discriminate to a higher degree a measurand from others of similar description, which means it supplies more information on the measurand. With this meaning of 'indifference ranges' the concept is suited to quantitative treatment and may turn out quite handy in tricky issues as the pattern recognition.

5. TIME AND MEASUREMENT

Time affects measurement as an influence quantity like the others, playing two different roles: (a) as the time allowed for the measurement operation; (b) as the lifetime of the instrument (or of its calibration).

The classical theory is ill equipped to cope with the first problem, as the true value is surely instantaneous (geometry is time independent) and no question of principle can arise with time-varying measurands. In practice a distinction is made between static and dynamic or stationary and transient measurements (definitions and terms change with the milieu); most standards refer only to the former, as no conceptual tools are available to deal with the latter. With an operative approach in terms of signals and assigned value-spans the measurement time fits snugly in the picture, as one of the influence quantities determining the width of the value-span, and one wonders whether a general relation might be worked out between uncertainty (as indifference range) and measurement time, similar to the one tying the indeterminacy of conjugate variables in quantum mechanics.

The second role refers to the time dependence of the calibration, hence to the problem of reliability. The cussedness of measurement depends on time [12] and a distinction is usually made between repeatability and stability, but the current approach does not go much further: the 'age' of a calibration is not usually considered an influence quantity though there is no reason for not treating it as such. When we calibrate an instrument we warrant the consistency of the value-spans assigned under the calibration; how long do we mean the warranty to hold? Not forever. Shall we state that the confidence level of the warranty decreases in time or that the uncertainty increases? The calibration is a procedure of quality control on the instruments, carried out with statistical techniques; the question then arises whether the ergodic theorem may be applied to a population of instruments to predict the behaviour of single ones [4, 15].

REFERENCES

[1] L. Gonella, Alta Frequenza, 4, 622 (1975).
[2] L. Gonella, IMEK0 TC7 Symposium, Enschede 1975.
[3] L. Gonella, seminar held at the Course "Metrology and Fundamental Constants", Int. School of Physics "E.Fermi", Varenna, July 1976.
[4] L. Gonella, IMEK0 TC7 Symposium, Leningrad 1978.
[5] H.V. Helmholtz, Zahlen und Messen erkenntnis – theoretisch betrachtet, Philosophische Aufsatze Eduard Zeller gewidmet, Leipzig 1887.
[6] S.S. Stevens, Measurement, Definitions and Theories (C.W. Churchman and P. Ratoosh Eds.), New York 1959, 20.
[7] B. Ellis, Basic Concepts of Measurement, Cambridge 1968.
[8] J. Pfanzagl, Theory of Measurement, Würzburg 1971.
[9] D.H. Krantz, R.D. Luce, P. Suppes, A. Tversky, Foundations of Measurement, Vol. 1, New York 1971.
[10] L. Finkelstein, Measurement and Control, 8, 105 (1975).
[11] B. Russell, The Principles of Mathematics, 2nd ed., New York 1937.
[12] C. Eisenhart, Journal of Research NBS, 67C, n°2, 161 (1963).
[13] L'Unificazione, 30, n°1, 21 (1976).
[14] J.L. Destouches, IMEKO TC7 Symposium, Enschede 1975.
[15] J.L. Destouches, IMEKO TC7 Symposium, Leningrad 1978.