Measurement: the limits of empirical knowledge

Luca Mari and Dario Petri

Measurement is a process producing trustworthy information about empirical properties of objects. This paper is devoted to describe why and how the information on the trustworthiness of the value or the values attributed to the property intended to be measured is provided. Some subjects of the paper are: specificity of measured values, meaningfulness and usefulness of measured values, measurement uncertainty.

Introduction

As presented and justified in the previous papers of this series about fundamentals of measurement [1-3], we are proposing to understand measurement as a process producing information about empirical properties of objects in such a way that the trustworthiness of the reported value or values attributed to the property intended to be measured, i.e., the measurand, is somehow explicitly stated and therefore publicly known. This condition explains the societal importance of measurements, whose results can be thus exploited, at a given degree of reliability, for decision making [4]: as such, measurement is a tool for intersubjectivity. How to obtain and to report the information on the trustworthiness of the reported value or values – hence actually a meta-information – is then a fundamental issue for measurement science, with critical consequences also for the daily practice in the many contexts in which measurement results are expected to enable fair agreements on the decisions to be made.

Trustworthiness is not a feature of values as such: entities like 0.123 kg or 1.234 m can only be more or less specific in the information they convey, so that for example 0.1234 kg is more specific than 0.123 kg, but they are not trustworthy, nor non-trustworthy, per se. Rather, trustworthiness is about how appropriately a given value is attributed to a given measurand. In the previous papers of this series we discussed what we consider the simplest understanding of this attribution, in terms of the relation

measurand = measured value

for example

mass of object A = 0.123 kg

interpreted, literally, as an equation claiming that the mass of object A is the same as the mass that is 0.123 times the mass that is the kilogram. This paper is devoted to discuss what limits the trustworthiness of this kind of claims and how to assess it.

Characterizing the Specificity of Measured Values

The background understanding of a relation like

mass of object A = 0.123 kg

interpreted as an equation is that entities like 0.123 kg *operate as classifiers*, so that the actual content of the relation is that

(1) we were able to construct a classification of masses through the kilogram and some of its multiples and submultiples, and

(2) the mass of object A belongs to the class identified as 0.123 times the kilogram

(note that this situation can be also interpreted in functional terms, by writing the relation above as

in_kilograms(mass of object A) = 0.123

where in_kilograms is a function – a scale in a metrological context – whose range establishes how masses are classified).

It is because of this that if the same value is attributed to the mass of different objects we are justified to infer that the two objects "have the same mass" – as it is customarily said – even if we never empirically compared them by mass: because the masses of the two objects belong to the same class in the relevant classification.

The crucial point here is that *more than one classification is possible for the same kind of property*, and the masses of two objects that belong to the same class in a rough classification could be revealed to be different in a finer classification. Hence, for any pair of classifications C_i, C_j for a given kind of property, if C_i is a refinement of C_j then the information provided by an element of C_i is more *specific* than the information provided by an element of C_j. In the case of ratio quantities, such as mass, some classifications can be simply characterized by the order of magnitude of the least significant digit (henceforth "OoM" for short) of the values identifying the elements of the classifications, as shown in Fig.1. Refining a classification, and therefore increasing the specificity of the produced information, corresponds in these cases to decreasing the OoM of the values. For example, reporting the value 0.1234 kg instead of 0.123 kg implies a refinement of the classification by one OoM, from 10^{-3} kg to 10^{-4} kg, where in fact, as already mentioned, 0.1234 kg conveys more specific information than 0.123 kg.

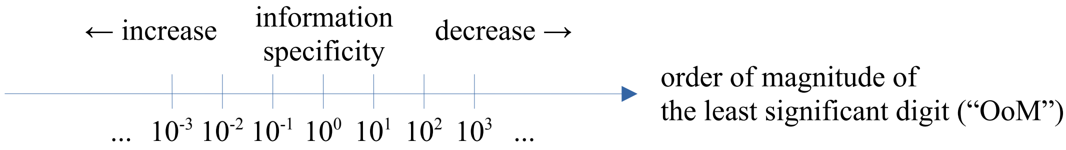

Fig.1. Specificity of classification based on Order of Magnitude (OoM).

We are thus adopting here what could be called an "operational interpretation of numbers": while mathematically integers are embedded in rationals and rationals are embedded in reals, so that for example 3, 3.0, and 3.000... is the same number, we are maintaining them distinct, given their different OoM.

The choice of the classification to use, and therefore of the possible values to report, is a matter of *appropriateness*, not truth. For example, the masses of planets could be appropriately classified with an OoM of 10^{20} kg, while the masses of humans with an OoM of 10^{-1} kg. Such appropriateness is about two complementary criteria.

The first criterion is related to both the kind of objects under consideration and the way their concerned property is defined, which *set the best specificity* for a classification to be able to convey *meaningful* information, corresponding to a lower bound of OoM, O_{low}. For example, the masses of humans may be classified with an OoM of 10^{-1} kg or possibly 10^{-2} kg, whereas classifying them with an OoM of, say, 10^{-6} kg is usually not meaningful, because the mass of a human body is not defined at that degree of specificity due to the state in which it is defined and to its changes over time (both the *Guide to the expression of uncertainty in measurement* (GUM) and the *International Vocabulary of Metrology* (VIM) treat O_{low} as an uncertainty, and call it "intrinsic uncertainty" [5] or "definitional uncertainty" [6] respectively).

The second criterion is related to the kind of use, including any decision making activity, expected for the produced information, which *sets the worst specificity* for a classification to be able to convey *useful* information, corresponding to an upper bound of OoM, O_{high}. For example, the masses of humans may be classified with an OoM of 10^{-1} kg or possibly 10^{0} kg, whereas classifying them with an OoM of, say, 10^{3} kg is not useful, because the mass of a human body is not discriminated at that degree of specificity (the VIM treats O_{high} as an uncertainty and calls it "target uncertainty" [6]).

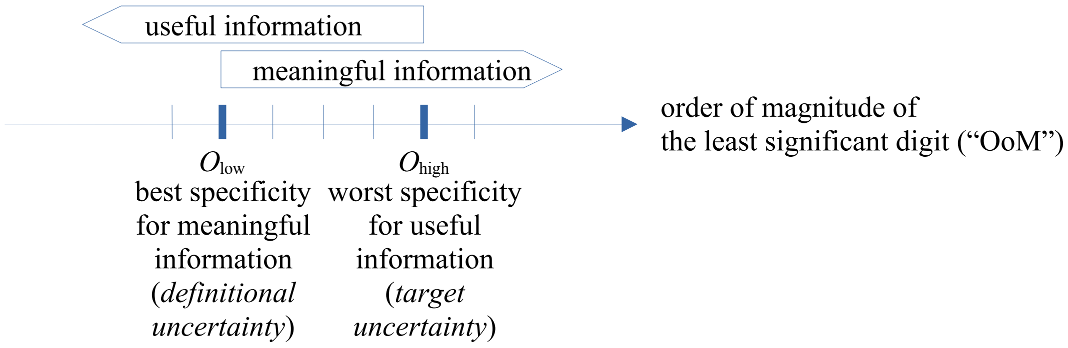

Fig.2. Range of appropriate specificity for classifications.

Hence, these two criteria taken together establish a range of appropriate specificity for classifications, such that the information provided by values of a quantity is

– meaningful if with an OoM not less than O_{low} (otherwise the information is not meaningful because too specific), and

– useful if with an OoM not greater than O_{high} (otherwise the information is not useful because not sufficiently specific).

Accordingly, as shown in Fig.2, we may call *best adequate classification* (relatively to the given definition of the measurand) and *worst adequate classification* (relatively to the given purpose of the measurement) the classifications with O_{low} and O_{high} respectively.

Under the plausible hypothesis that the resources required to produce trustworthy information on the property of a given object through its evaluation increase monotonically with the increase of the degree of specificity of the produced information, this range also provides some pragmatic guidelines:

– values more specific than O_{low} (e.g., reporting values of mass of humans with an OoM of the milligrams) convey information that is only partially meaningful, and therefore that was produced by wasting some resources;

– values less specific than O_{high} (e.g., reporting values of mass of humans with an OoM of the tonnes) convey information that is not useful, and therefore that was produced by wasting the used resources.

Of course, if $O_{low} > O_{high}$ (e.g., the masses of humans are acknowledged to be classifiable with an OoM of 10^{-1} kg but information about them would be required with an OoM of 10^{-2} kg) then producing meaningful and useful information by property evaluation is in principle not possible.

Vice versa, $O_{low} \leq O_{high}$ is the sort of situations that justify devoting some resources to produce trustworthy information on the concerned property.

Values of Properties as Classifiers

Empirical properties of the same kind can be compared with each other, at least by indistinguishability and sometimes also by order, or difference, or ratio, or number. Such a comparison is, generally, *not* adequately formalized as an equivalence relation, and therefore what results from it is, generally, *not* a classification. Indeed, for a relation to be an equivalence it must be reflexive, symmetric, and transitive. While reflexivity is trivial (the mass of any object A, M_A, is indistinguishable from itself, $M_A \approx M_A$) and symmetry is not an issue (for any pair of objects A and B, if $M_A \approx M_B$ then $M_B \approx M_A$, i.e., the order with which they are compared is not important), transitivity is generally not guaranteed: for some triples of objects A, B, and C, it may well happen that $M_A \approx M_B$ and $M_B \approx M_C$, and nevertheless not $M_A \approx M_C$, i.e., not necessarily empirical indistinguishability chains.

The point is that without transitivity the comparison does not lead to distinct classes of properties, but only to mutually overlapping subsets (in the example above about masses, A and B belong to one subset and B and C to another subset, but there is not a subset to which both A and C belong). This is critical, because in this situation the properties of objects can be identified only individually, as the classical paradox known as "sorites" shows: for example, while a heap of sand from which a single grain is removed remains a heap, by applying repeatedly this process the situation is obtained that any set of grains is a heap. To apply the paradox to our case it is sufficient to substitute 'to be a heap' with 'to have indistinguishable mass'.

The messy situation of non-transitive empirical indistinguishability between properties of object, like $M_A \approx M_B$, is solved by relying on values of properties, which are ultimately expected to be connected to empirical properties via metrological traceability chains: the scientific, technological, organizational, and political infrastructures aimed at guaranteeing that, say, one kilogram is the same mass all over the world [7]. Under the condition that values are traceable to one and the same unit, the relation

$R(M_A, M_B)$ if there is a measured value v such that $M_A = \mathrm{v}$ and $M_B = \mathrm{v}$

is an equivalence. This is a key reason for dealing with values of properties: because, as stated, they are information entities that operate as classifiers for properties of objects.

True Value: an Operational Perspective

Even before dealing with the quality of the information produced by a property evaluation, thus in particular a measurement, what is described above allows us to ponder about one of the most controversial concepts of metrology: the idea that empirical properties may have a true value (for a recent presentation and analysis of this subject see [8]). Around true values there are several misunderstandings in metrology, one of them is well presented so: "The input to the measurement system is the true value of the variable; the system output is the measured value of the variable. In an ideal measurement system, the measured value would be equal to the true value." [9]. This position, according to which an ideal measurement realizes an identity function (and as such would be useless?), is so abstract – as the use of the term "variable" suggests, by the way – that it neglects the basic distinction between

– properties known empirically, as properties of objects: the input of measurement, and
– properties known informationally, as values of properties: the output of measurement.

Furthermore, this position, which echoes the traditional adage "the true value is what would be obtained when all measurement errors are removed", assumes that true values are independent of classifications, scales, or OoMs. Rather, from the consideration that values of properties are identifiers for elements of a classification we can derive *an operational concept of true value*, to be intended as 'true value in a given classification' (or 'true value in a given scale'), where the OoM is then a priori set.

Let us assume that the OoM of the values to be reported is not less than O_{low}, so that there is one class in the chosen classification to which the measurand belongs, as it could be, for example, if the mass of a given person A were classified with an OoM of 10^{-1} kg or 10^{0} kg. In this situation the equation

mass of person A = x kg

is either true or false for any given number x with one decimal digit, and it is true for the value x kg that identifies the class to which the measurand belongs. And in this situation the value x kg for which the equation is true can be operationally considered to be *the true value of the measurand with respect to the given classification*: the mass of any given person, in a given moment, has a true value in kilograms, and so on.

Of course, the fact that a concept is operationally well defined does not imply that we are able to assess if it applies to any given empirical situation. In our case, the less specific is the classification, the more probable becomes the claim that a value provided by measurement is the true value of the measurand: for example, in the scale of 10^{1} kg we are usually sure of our mass (the Appendix discusses the opposite case, in which the OoM of the values to be reported is less than O_{low}, as when numerical values are real numbers with infinitely many significant digits, the situation that allows us to apply the tools of calculus).

Measurement Uncertainty

Calibrated measuring instruments are designed to produce measured values by operationalizing the equation

measurand = measured value

in functional terms

f(measurand) = measured value

where the function *f*, what in [3] we called a *Measurement Mapping*, describes the process of measuring the measurand. Given the position that property values operate as classifiers, the obvious consequence is that, if it actually produces a single measured value, *measurement is a classification process*, and performing a measurement induces in principle a classification over the set of the possible measurands. This classification – let us call it a *measurement classification* –, which establishes the set of possible measured values, depends on the metrological features of the adopted measuring system, first of all its resolution, i.e., the "smallest change in a quantity being measured that causes a perceptible change in the corresponding indication" [6] (def. 4.14). For the sake of simplicity, let us suppose that the classification induced by the measuring system is characterized by the order of magnitude of the least significant digit of the measured values, designated as O_{meas}, which typically corresponds to the resolution of the displaying device. This is the "smallest difference between displayed indications that can be meaningfully distinguished" for instruments with analogue reading ([6] def. 4.15)) and to the least significant digit in the display for instruments with digital reading. The consideration above provides us with the now straightforward summary conditions:

– if $O_{meas} < O_{low}$ (i.e., the measurement classification is more refined than the best adequate classification, as if the mass of humans were measured with a specificity of the milligrams), the resolution of the measuring system is higher than required, so that part of the information it produces is not meaningful: a less sophisticated, and then plausibly cheaper, measuring system would have been sufficient;
– if $O_{meas} > O_{high}$ (i.e., the measurement classification is less refined than the worst adequate classification, as if the mass of humans were measured with a specificity of the tonnes), the resolution of the measuring system is lower than required, so that the information it produces is

not useful: a more sophisticated, and then plausibly more expensive, measuring system would have been necessary.

Accordingly, once O_{low} and O_{high} are established, by defining the measurand and deciding the minimum specificity of the information to be produced respectively, the condition

$$O_{\mathrm{low}} \leq O_{\mathrm{meas}} \leq O_{\mathrm{high}}$$

is a precondition for a well-designed measurement setup (see Fig.3).

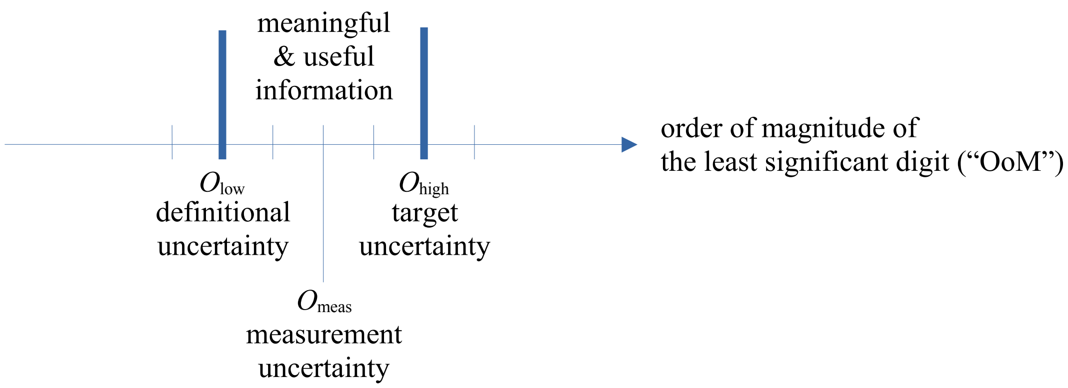

Fig.3. Constraints on measurement uncertainty ensuring a well-designed measurement setup.

As a matter of fact, various causes, among those that the GUM calls “sources of uncertainty” [5] (item 3.3.2), may prevent to perform a measurement that reliably chooses the element of the classification whose OoM is O_{meas} that corresponds to the true value of the measurand, This leads us to acknowledge that what a measurement produces is usually not one such measured value, but something able to convey information on both one or more values attributed to the measurand and the trustworthiness of such attribution. Several options are possible to this purpose, all of them informally describable as cases of *measurement uncertainty*, and thus generalizing the GUM definition, “parameter, associated with the result of a measurement, that characterizes the dispersion of the values that could reasonably be attributed to the measurand” [5]. Let us shortly discuss some of them.

– A measurement result could be one value whose OoM is greater than O_{low}, hence in this case the measurement classification is less refined than the best adequate classification. Suppose, for example, that the mass of a given person A is considered knowable in the scale of 10^{-1} kg (the O_{low}) but the available measuring system has a resolution of the kilogram (the O_{meas}) and produced 65 kg as the measured value. In this case the values 64.5 kg, 64.6 kg, ..., 65.5 kg, as identifiers of the elements of the best available classification, are all compatible with the measured value, and the range of that interval implicitly provides the information on the measurement uncertainty (thus also implying that, say, 65.4 kg, cannot be reported as the measured value, since it is below the threshold of specificity of the information obtainable by the given measuring system).

– A measurement result could be one value related to the measurement classification together with a scale statistic. It is the basic option of the GUM, that assumes the measurand to be modeled as a random variable, whose standard deviation is estimated by statistical means or otherwise ("Type A" and "Type B" evaluation of measurement uncertainty respectively [6]). Such an estimate – called a *standard uncertainty* – can be then the standard deviation of the mean of the available sample, under the hypothesis that the value of the measurand is estimated by the sample mean.
– A measurement result could be a set of values related to the measurement classification that could be attributed to the measurand, possibly together with an explicitly or implicitly specified coverage probability. It is the option suggested by the GUM when the measurement result is employed to support decision making, and in particular conformity assessment [10]. According to the GUM, that set of values is defined through the so-called *expanded uncertainty*, obtained by multiplying the measurement standard uncertainty by a coverage factor (usually between 2 and 3), related to the coverage probability associated to the measurement result. Of course, this requires some assumptions on the underlying probability distribution over the whole set of the values that might be returned by the measurement classification.
– A measurement result could be finally a probability distribution over the whole set of the values returned by the measurement classification. In this case measurement uncertainty has to do with the shape of the distribution. To provide a concise result and to simplify subsequent processing, the information associated to the probability distribution may be summarized by a scale parameter or a set of values, thus falling in one of the two previous options.

A measurement system can be a complex entity, whose behavior results from

– several "low-level" features, like sensitivity, selectivity, stability, and resolution [6], and
– two "high-level" features, precision and trueness, together contributing to accuracy [11].

With the aim of effectively exploiting measurement information for prediction or decision making, users need a socially harmonized way to deal with such a complexity. In the past this was the role of *measurement error*, which focuses on the distance from the ideality that affect any empirical process. The developments of metrology in the last decades have led to a broader standpoint, in which measurement errors are only a cause, though possibly an important one, of the less-than-requested objectivity and intersubjectivity of measurement results. Measurement uncertainty has become *the* tool to provide information on the limits of the empirical knowledge, and therefore to summarize the trustworthiness of the information produced by measurement.

Appendix: On the Existence of True Values

Complementary to what discussed in the section “True value: an operational perspective” above, let us assume here that the OoM of the values to be reported is less than O_{low}, as in the extreme case in which numerical values are real numbers with infinitely many significant digits, so that something like π kg might even be reported as a value of mass, as obtained by computation and surely not by any empirical means. It is a fact that all digits below O_{low} are not meaningful in this situation, i.e., they do not provide information about the measurand. We envisage two possible alternative interpretations on this matter.

First, in such a situation *no value can be true*, so that the very concept ‘true value’ loses its possibility of a reference. Coming from decades in which true values have been considered as non-existing entities, it is plausible that this interpretation would be accepted by those who endorse a representational and at least mild relativist standpoint.

Second, in such a situation *a measurand has multiple true values*. Let us suppose, for example, that 65.4 kg is the value that identifies the class to which the measurand belongs (how this value has been obtained is not the issue here: again, we are discussing about the concept ‘true value’ and the possibility of existence of true values of measurands, not our ability to know them). Then, any value 65.4z kg, where z is a sequence of one or more digits (hence any value like 65.41 kg, 65.411 kg, …, 65.42 kg, …) identifies an empirically undefined subclass of the class to which the measurand belongs. While the increased specificity provided by the digits z in the numerical value is not meaningful, all these values may be considered true values of such a measurand. This interpretation seems to be consistent with the VIM, where it notes (def. 2.11, Note 3) that “when the definitional uncertainty associated with the measurand is considered to be negligible compared to the other components of the measurement uncertainty, the measurand might be considered to have an “essentially unique” true quantity value” [6]. When instead definitional uncertainty is not negligible, O_{low}, i.e., definitional uncertainty itself, sets the range of the interval of true values of the concerned measurand.

References

[1] L. Mari and D. Petri, "Fundamentals of measurement: knowledge from information about empirical properties", *IEEE Instr. Meas. Mag.*, vol. 26, no. 1, 2023.

[2] L. Mari and D. Petri, "Fundamentals of measurement: the social spread of knowledge", *IEEE Instr. Meas. Mag.*, vol. 26, no. 4, 2023.

[3] L. Mari and D. Petri, "Fundamentals of measurement: designing the acquisition of knowledge", to be published in IEEE Instr. Meas. Mag.

[4] D. Petri, P. Carbone, and L. Mari, "Quality of measurement information in decision-making", *IEEE Trans. Instrum. Meas*., vol. 70, pp. 1–16, 2021.

[5] *JCGM 100, Evaluation of measurement data – Guide to the expression of uncertainty in measurement (GUM)*, Joint Committee for Guides in Metrology, 2008. [Online]. Available: https://www.bipm.org/documents/20126/2071204/JCGM_100_2008_E.pdf.

[6] *JCGM 200, International Vocabulary of Metrology – Basic and general concepts and associated terms (VIM)*, 3rd ed., Joint Committee for Guides in Metrology, 2012. [Online]. Available: https://www.bipm.org/documents/20126/2071204/JCGM_200_2012.pdf.

[7] L. Mari, S. Sartori, "A relational theory of measurement: traceability as a solution to the non-transitivity of measurement results", *Measurement*, vol. 40, no. 2, pp. 233-242, 2007.

[8] F. Grégis, "Do quantities have unique true values? The problem of non-uniqueness in measurement", *Measurement*, vol. 221, 2023

[9] J. P. Bentley, *Principles of measurement systems*. 4th ed., Pearson, 2005.

[10] JCGM 106, *Evaluation of measurement data – The role of measurement uncertainty in conformity assessment*, 2012. [Online]. Available: https://www.bipm.org/documents/20126/2071204/JCGM_106_2012_E.pdf.

[11] *ISO 5725-1, Accuracy (trueness and precision) of measurement methods and results – Part 1: General principles and definitions*, International Organization for Standardization, 2023.

Possible pull quotes:

- In appropriate conditions of specificity, the true value of a measurand can operationally be defined as corresponding to the class to which the measurand belongs
- Various causes may prevent a measurement to identify the class to which the measurand belongs: measurement uncertainty quantifies the trustworthiness of the provided classification, and as such it is an essential component of measurement result

Luca Mari is currently a Full Professor in Measurement Science with Università Cattaneo - LIUC, Castellanza, Italy, where he teaches courses about measurement science, statistical data analysis, and systems theory. He received the M.Sc. degree in physics from the University of Milano, Italy, in 1987, and the Ph.D. degree in measurement science from the Polytechnic of Torino, Italy, 1993. Dr. Mari has been the Chair of the TC7 (Measurement Science) of the International Measurement Confederation (IMEKO) and the Chair of the TC1 (Terminology) and the Secretary of the TC25 (Quantities and units) of the International Electrotechnical Commission (IEC). He is an IEC Expert in the WG2 (VIM) of the Joint Committee for Guides in Metrology (JCGM).

Dario Petri is currently a Full Professor in Measurement Science and Electronic Instrumentation at the Department of Industrial Engineering of the University of Trento, Italy. He received the M.Sc. and Ph.D. degrees in electronics engineering from the University of Padua, Italy, in 1986 and 1990, respectively. He is an IEEE Fellow member and the recipient of the 2020 IEEE Joseph F. Keithley Award for "contributions to measurement fundamentals and signal processing techniques in instrumentation and measurement." Dr. Petri's research activities are focused on digital signal processing applied to measurement problems, data acquisition systems, and fundamentals of measurement science.