A structural framework across strongly and weakly defined measurements

Luca Mari
School of Industrial Engineering
Università Cattaneo – LIUC
Castellanza (VA), Italy
lmari@liuc.it

Mark Wilson
Graduate School of Education
University of California
Berkeley, CA, USA
markw@berkeley.edu

Abstract—Is there a framework common to measurement across broad domains of application, including both physical and social science measurement? The answer to this question would determine to an important extent the possibility of building a shared measurement-related body of knowledge across many traditionally separate domains. In this paper, we outline a structural framework of the processes involved in the construction and use of measurement that includes instrument design, standard/reference selection, instrument calibration, measurand definition, and ultimately, instrument operation and data processing.

Keywords—fundamentals of measurement; strongly and weakly defined measurement; physical and social measurement

I. INTRODUCTION

The widespread adoption of measurement processes in many domains has produced a multiplicity of conceptualizations and lexicons even for the basic entities that are expected to be constitutive of a measurement. An interesting characterization on this matter was proposed by Ludwik Finkelstein in terms of strongly defined and weakly defined measurement: “Strongly defined measurement [...] is based on (i) precisely defined empirical operations, (ii) mapping on the real number line on which an operation of addition is defined, (iii) well-formed theories for broad domains of knowledge. [...] Weakly defined measurement is based on an objective empirical process, but lacks some, or all, of the above distinctive characteristics of strong measurement” [1, p.42].

With the aim of introducing what we hope will constitute a broadly-sharable conceptual framework encompassing both strongly and weakly defined measurement, and therefore including both physical and social science measurement, we propose a structural view of the measurement process, in which we analytically identify components of the process while taking into account only the function performed by each component, not the way it is constructed. Components are described as inter-connected through their input-output relations, thus giving a systemic interpretation of the process. Such a structural view makes the presentation independent of the level of quality (from rough to highly accurate) of modeled processes, of the evaluation type they realize (from nominal to absolute, i.e., from classification to counting), and of the nature of the measurands (physical or non-physical). Our intention is that this focus on structure will result in a harmonization of viewpoints across physical and social science contexts.

The core idea behind our proposal for this framework is not so controversial nowadays: measurement is more than a purely experimental process. Indeed, the experimental activity of making a measuring instrument interact with an object under measurement actually produces measurement results only if some preconditions, both conceptual and experimental, are satisfied [2]. Less obvious perhaps is the specification of these preconditions, their functional role, and their relations: an instrument must be appropriately designed; a reference set must be appropriately selected; the instrument must be calibrated with respect to the reference set; the measurand must be defined. On this basis, the framework gives an explanation of the nature, the source, the features, and the limits of the information conveyed by measurement.

II. JUSTIFICATION

Is a structural view of measurement appropriate? Is not it just a re-proposition of operationalism? (“In general, we mean by any concept nothing more than a set of operations; the concept is synonymous with the corresponding set of operations.” [3, p.5]). Should not measurement be defined in reference to the kind of information it provides, or the kind of entities it properly applies to? [4]

Such objections are interesting, and have to be taken seriously given the flaws acknowledged in radical operationalism, in its claim that any concept should be defined in terms of operations.

The distinction between what we are discussing here and the perspective from operationalism is that we are not claiming that the framework defines the measurand, as Bridgman would have done (e.g., “The concept of length is therefore fixed when the operations by which length is measured are fixed: that is, the concept of length involves as much as and nothing more than the set of operations by which length is determined.”). Instead, the framework specifies that the measurand definition is a component of the process, hence the framework cannot define the measurand, nor any other property involved in the measurement process. Rather, even if the framework is considered as a definitional device (instead of as a tool to support the development of a measurement), what it defines is the concept of measurement itself: measurement is a property
evaluation process performed according to this structure, where the superordinate concept ‘property evaluation process’ is assumed to be previously and independently defined (plausibly in the context of a background ontology including properties and property values as basic entities).

While defining a property by means of operations is a reductionist strategy, there is nothing reductionist in defining a process through an operational structure. This avoids the many stereotypes still cluttered around the concept ‘measurement’ and ‘measure’, and in this sense is consistent with the neo-positivist dictum: “a scientific description can contain only the structure (form of order) of objects: not their ‘essence’” [5]. On the other hand, precisely in order to reduce the risk of paradigm incommensurability [6], the framework will be introduced as characterizing a generic process that we call experimental property evaluation, of which we claim measurement is a specific case. Whether any experimental property evaluation is a measurement, or the framework only provides a necessary condition for measurement, might remain a purely semantic issue.

What is not semantic, and surely not conventional, is instead the social trust attributed to measurement results, as distinct from the case for subjective opinions: a conceptual framework on measurement must be then a necessary part of this trust, that cannot rely only on the traditional hypothesis that measurement produces quantitative information, given that also opinions can be stated in quantitative terms.

Our claim is that measurement is a process whose structure is designed so as to enable the production of information that is simultaneously (a) about the intended object of measurement and (b) equally interpretable by different subjects, i.e., information that is objective and inter-subjective [7].

Let us assume a background ontology such that there are:

- objects \( a \) (objects can be bodies, events, processes, persons, organizations, etc); objects are time persistent, i.e., \( a(t) \) is the object \( a \) at the time \( t \);
- general properties \( X \) (length, attitude, etc);
- individual properties \( X(a) \) (the length of a given body, the attitude of a given person, etc), or \( X(a(t)) \) when the reference to time is taken into account (the length of a given body at the time \( t \), etc).

Each individual property is the instance of a general property (the length of this body is a length, etc), and a general property can be characterized by the possibility that its instances to be compared with each other, the simplest case of comparison being just equality or difference, while cases more complex for the implied algebraic structure are order, distance, and ratio [8]. At least in some basic cases the assessment of difference or equality (possibly in the empirical sense of: “no observed difference” vs “observed difference”) can be performed as a primitive, pre-theoretical operation, thus in particular independently of any formal / mathematical apparatus.

The process of measurement is ultimately aimed at reporting information on an individual property, where the concept ‘information’ is meant here in its technical, syntactical dimension of recognition of differences: “it is this, while it might have been that” [9]. This further justifies characterizing measurement within the encompassing frame of a property evaluation processes.

III. STRUCTURAL DESCRIPTION OF AN EXPERIMENTAL PROPERTY EVALUATION

Measurement is a specific case of a process that produces information on an individual property as a value (“one or more values” according to the International Vocabulary of Metrology [10], to keep into account uncertainties), a “property evaluation” for short [7]. By proposing a structural description of experimental property evaluation, we aim at introducing a framework that, at the same time, is able to:

- identify certain structural analogies between measurement, as traditionally understood, and experimental property evaluation: this may lead to the extension of the concept of measurement through sufficient conditions that are weaker than sometimes assumed;
- identify certain structural conditions required for a property evaluation to be considered a measurement: this may lead to the restriction of the concept of measurement through necessary conditions that are stricter than sometimes assumed.

These changes would be ultimately justified by structural consistency.

In what follows we exemplify the context of measurement as the process to produce information as temperature values from the experimental operation (hence an experimental property evaluation) of, say, a thermometer based on a thermocouple sensor, i.e., a temperature-electric potential difference (voltage) transducer.

A. Instrument design

One observes that a property \( Y \) of an object \( b \) changes (say, the voltage \( Y(b) \) of \( b \), a device made by two different wires joined at one end), and from further observations she hypothesizes that the changes of \( Y(b) \) are systematically and regularly caused by the changes of a property \( X \) of other objects \( a_i \) with which \( b \) somehow interacts (say, the temperature \( X(a_i) \) of \( a_i \), the environment around the junction end of \( b \)). “Systematically and regularly” only means here that, ceteris paribus, if at different times, \( t_1 \) and \( t_2 \), one observes \( Y(b(t_1)) \) is [not] equal to \( Y(b(t_2)) \) (the voltage of \( b \) is [not] the same at the times \( t_1 \) and \( t_2 \), then she hypothesizes that also \( X(a(t_1)) \) is [not] equal to \( X(a(t_2)) \) (the temperature of \( a \) is [not] the same at the times \( t_1 \) and \( t_2 \)).

This inference crucially assumes that:

- the properties \( Y(b(t_i)) \) can be somehow stored;
- a stored property \( Y(b(t_i)) \) can be compared with a subsequently obtained property \( Y(b(t_j)) \), \( t_j > t_i \) (how this can be obtained, e.g., by means of stable marks on a stable scale, is not discussed here).
In this situation one can write \( F_x(X(a)) \) to denote the property \( Y(b) \) caused by \( X(a) \), where then \( F_x \) is a causation mapping that describes the transduction \( X \rightarrow Y \) performed by \( b \) (hence, for the configuration of wires \( b \), the mapping is \( F_x \): temperature \( \rightarrow \) voltage, where \( Y(b) = F_x(X(a)) \) means: voltage at the junction end of \( b = F_x \) (temperature in a point a close enough to the junction end of \( b \)).

One also observes that the same behavior of \( b \) is obtained by other objects \( b_j \), structurally analogous to \( b \), through their interaction with other objects \( a_j \), i.e., all \( a_j \) have the property \( X \) and all \( X(a_j) \) systematically and regularly produce corresponding changes of \( Y(b_j) \) (i.e., different analogous configurations of wires \( b_j \) produce this transduction temperature-voltage behavior when applied in different points \( a_j \), eventually leading to the (re)discovery of the Seebeck effect, being thus each \( b_j \) a thermocouple).

Let us call an object \( b_j \), when used as above, a transducer (or also, more specifically, a sensor, “element of a measuring system that is directly affected by a phenomenon, body, or substance carrying a quantity to be measured”, according to [10]), and let us call \( X(a) \) its input property and \( Y(b) = F_x(X(a)) \) its output property (or indication).

In general, different transducers \( b_1, b_2, \ldots \) can produce different transduction results, \( F_{b_1}(X(a)) \neq F_{b_2}(X(a)) \neq \ldots \), due to differences in their internal structure. Moreover, the output property \( Y(b) \) generally depends not only on the input property \( X(a) \), but also on some other properties \( W_m(c) \), where \( c \) can be \( a \), or \( b \), or their environment, i.e., \( Y(b) = F_x(X(a), W_l(c), W_2(c), \ldots) \) (we use the simpler notation \( F_x(X(a)) \) when the dependence on \( W_m(c) \) is not relevant). Let us call each \( W_m(c) \) an influence property for the transducer.

Three basic conditions are then expected from the behavior of a transducer \( b \):

- if \( X(a_1) \) and \( X(a_2) \) are “sufficiently different”, then \( F_x(X(a_1)) \neq F_x(X(a_2)) \) (X-sensitivity);
- if \( X(a(t_1)) \) and \( X(a(t_2)) \) are not “too different”, then \( F_x(X(a(t_1))) = F_x(X(a(t_2))) \) (time-insensitivity, also called X-stability);
- if \( X(a_1) \) and \( X(a_2) \) are “sufficiently equal” and \( W_m(c_1) \) and \( W_m(c_2) \) are not “too different”, then \( F_x(X(a_1), W_m(c_1)) = F_x(X(a_2), W_m(c_2)) \) (W-m-insensitivity, also called W-m-selectivity).

A transducer whose behavior fulfills these conditions produces an output property that sufficiently depends exclusively on the input property: under the supposition that the transducer is expected to be only sensitive to its input property, this is a meta-condition of object-dependence, i.e., objectivity.

When it is observed or hypothesized that the transducer \( b \) does not behave exactly as expected, this reduces the objectivity of the process and is a source of uncertainty in its outcomes.

**B. Standard/reference selection**

Suppose a set of objects \( a_k \) is available, \( k > 0 \), let us call them standards, such that:

- each \( a_k \) has the property \( X \), specifically, \( X(a_k) \), and \( X(a_1) \neq X(a_2) \neq \ldots \) (e.g., \( a_k \) is a sample of freezing water, \( a_2 \) is a sample of boiling water, etc);
- they can be put in interaction with transducers \( b \), thus producing \( F_x(X(a_k)) \);
- they are socially widespread enough, so that the interaction of each of them with \( b \) can be easily obtained;
- they are X-stable enough (i.e., they are sufficiently time-insensitive), so that \( X(a_k(t_1)) = X(a_k(t_2)) \) where \( t_1 \) and \( t_2 \) are sufficiently distant (note that X-stability for transducers is a process stability, whereas for standards it is an object stability; transducers have inputs, standards do not).

Moreover, changes of \( X(a_k) \) could depend on changes of some other properties \( Z_n(c) \), where \( c \) is \( a_k \) itself or its environment, i.e., \( X(a_k) = X(a_k, Z_n(c), Z_n(c), \ldots) \) (note that the set of \( W_m \) above and the set of \( Z_n \) here need not be the same). In this case one more condition is required on the standards \( a_k \), i.e., that:

- they are \( Z_n \)-selective enough (i.e., they are sufficiently \( Z_n \)-insensitive), so that \( X(a_k, Z_n(c(t_1))) = X(a_k, Z_n(c(t_2))) \) where \( Z_n(c(t_1)) \) and \( Z_n(c(t_2)) \) are not “too different”.

Let us call each \( X(a_k) \) a reference property (\( a_k \) might be water in its freezing point, possibly under given pressure conditions, and \( X(a_k) \) its freezing temperature). Hence reference properties are properties of standards.

When it is observed or hypothesized that the standards \( a_k \) do not behave exactly as expected, this is a source of uncertainty in the properties they realize.

**C. Instrument calibration**

The property \( Y(b) \) is supposed to be causally dependent on the property \( X(a) \), which can be inferred from \( Y(b) = F_x(X(a)) \), according to the logic \( X(a) = F_x^{-1}(Y(b)) \). On the other hand, to be inverted, the causal mapping \( F_x \) must be known. Moreover, \( F_x \) depends on the specific transducer \( b \), and therefore knowledge about the generic transduction effect \( X \rightarrow Y \) is not enough: the behavior of the specific instrument \( b \) has to be characterized.

For a given transducer \( b \) of interest, a given set of reference properties \( \{X(a_k)\} \), and a time \( t_0 \), the transduced reference properties \( F_x(X(a_k)(t_0))) \), \( F_x(X(a_k)(t_0))) \), ... are obtained and stored (e.g., as distinct angular positions of the needle of a second transducer; mapping voltages to angular positions; each position can be marked by an identifier of the temperature \( X(a_k) \) that produced that position).

Let us call calibration the process that from the interaction of \( b \) with \( a_k(t_0) \) leads to the storage of \( F_x(X(a_k(t_0))) \). This
defines the transduction function $F_b$, for the arguments $X(a_1^*)$, as it can be represented by constructing a calibration table:

<table>
<thead>
<tr>
<th>$X$</th>
<th>$Y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X(a_1^*)$</td>
<td>$F_b(X(a_1^*))$</td>
</tr>
<tr>
<td>$X(a_2^*)$</td>
<td>$F_b(X(a_2^*))$</td>
</tr>
</tbody>
</table>

in which each reference property of $X$ in the first column is the entry to which the corresponding transduced property is associated in the second column (reference temperatures are then mapped to voltages of the thermocouple under calibration).

Calibrating a transducer makes it possible to compare each $F_b(X(a_1^*)(t_0)))$ with subsequent $F_b(X(a(t)))$, $t > t_0$: under the supposition that in the interval $[t_0, t]$ the transducer remained sufficiently $X$-stable and the standards remained sufficiently $X$-stable and $Z$-selective, this is a meta-condition of subject-independence, i.e., inter-subjectivity.

When it is observed or hypothesized that the calibrated transducer $b$ does not behave exactly as expected, this reduces the inter-subjectivity of the process and is a source of uncertainty in its outcomes.

**D. Measurand definition**

(Lexical note: the term “measurand” specifically refers to measurement, whereas we are discussing the more general case of property evaluation, and therefore we should adopt a different term, conveying a more generic meaning: “examinand” has been proposed [11], but there is a lack of terminological consensus we will stick to “measurand”, but intend it in a generic sense.)

By definition, the information obtained by measurement is referred to the measurand, i.e., the property “intended to be measured” [10]. Hence, in the specific and important case of the equation:

$$Q = [Q] \cdot [Q]$$

i.e., individual quantity = numerical value · unit [12], [13], the left hand term $Q$ denotes the measurand.

The transducer $b$ is supposed to interact with $a$ through its property $X(a)$, and therefore the usually implicit assumption is simply that $X(a)$ is the measurand. This is in fact a matter of the way the measurand is defined. In the simplest case the definition is entirely operational: the measurand is defined, literally, as the property with which the transducer interacts here and now. This makes the definition unproblematic (the transducer is designed, built, setup, and operated so to be able to interact with properties) but only of very limited usefulness: measurement results are generally produced to support object-related, and not instrument-dependent, decision making. Even that the property with which the transducer interacts here and now is a property of the object under measurement is an assumption based on a model of the interaction and a model of the measurand (consider, e.g., the difference between ‘temperature of this point of this object’ and ‘temperature of this object’).

When it is observed or hypothesized that these models are not perfectly realistic, this is a source of uncertainty in the measurand as reported in the measurement result.

**E. Instrument operation and data processing**

Given a transducer $b$ that has been calibrated in reference to a reference set $\{X(a_1^*)\}$, the relation ‘having the same $X$ as’ between $a$ and $a_1^*$, for a given $k$ (e.g., $a$ has the same temperature as freezing water), can be experimentally determined by making $b$ interact with each $a_k$, obtaining the transduced property $F_b(X(a_k))$ and then deciding according to the procedure:

$[P]$ if $F_b(X(a)) = F_b(X(a_1^*))$

then $X(a) = X(a_1^*)$ (the $X$ of $a$ and $a_1^*$ is the same)

else if $F_b(X(a)) = F_b(X(a_2^*))$

then $X(a) = X(a_2^*)$ (the $X$ of $a$ and $a_2^*$ is the same)

... (if the thermocouple produces the same voltage when applied to $a$ and to a sample of freezing water $a^*$ then $a$ and $a^*$ have the same temperature, else, etc.) (note that this procedure does not require the comparability, of $X(a)$ and $X(a_1^*)$, and instead requires only that $F_b(X(a_1^*))$ has been stored).

The calibration table is used here inversely, and the obtained transduced property in the second column is the entry from which the corresponding reference property in the first column is obtained as result. On the other hand, presenting the acquired information as $X(a) = X(a_k^*)$ for a given $k$, is not completely adequate, since it leaves it implicit the available information that $X(a) = X(a_k^*)$ for all $k \neq k’$. The result is then better reported as a selector in the reference set $\{X(a_1^*)\}$, i.e.:

$$X(a) = X(a_k^*) \in \{X(a_1^*)\}$$

where the reference set can be given a name, thus obtaining the usual way to report the results of nominal and ordinal property evaluations (e.g., the wind strength is here now $a_1^*$ in the Beaufort scale, the hardness of this sample is $a_1^*$ in the Mohs scale, etc.). This is the structurally simplest information that a property evaluation process can convey, i.e., $X(a)$ is recognized as equal to one and only one $X(a_k^*)$ (the chosen standard might be such that no $X(a_k^*)$ has been found that $F_b(X(a)) = F_b(X(a_k^*))$; in this case the set could be extended with new elements, but this situation is not further discussed here).

Fundamental here is that, while $F_b(X(a_1^*))$ and $F_b(X(a))$ depend on the transducer $b$ (they are indeed properties of $b$), the outcome of the procedure $[P]$ does not: in principle calibration has made the outcome independent of the transducer (while the voltage $F_b(X(a))$ generated by the temperature $X(a)$ depends on the thermocouple, that such temperature is either equal to or different from a reference temperature is independent of the thermocouple itself).

The several sources of uncertainty, related at least to the transducer behavior, the standards behavior, the transducer calibration, the interaction model, and the measurand model, contribute to the uncertainty of the reported result.
By adopting the usual terminology of measurement systems, then:

- \( a \) is the object under measurement (a point in space);
- \( a_k \) is an object used as a measurement standard (a sample of freezing water);
- \( b \) is used as a measuring instrument (a thermocouple);
- \( X \) is the property subject to measurement (temperature);
- \( Y \) is the instrument indication (voltage);
- \( X(a_k) \) is the property realized by the measurement standard (temperature of freezing water),

and:

- instrument operation and data processing corresponds to a measurement performed by means of \( b \) according to the procedure \([P]\).

Note also that:

- only the causal relationship between \( X \) and \( Y \), as realized by \( b \), is assumed: no analytical models are required;
- despite this simplicity, model-ladness is unavoidable, at least in the assumptions of \textit{ceteris paribus} comparison (i.e., all influence properties did not change), and stability of the transducer and the standard.

IV. CONCLUSIONS

The framework we have proposed provides a structure for the whole process of measurement, and more generally, experimental property evaluation, as constituted of two main stages, as follows.

- One stage aimed at satisfying the general preconditions for measurement (instrument design, standard/reference selection, instrument calibration); in this stage the measuring system is developed, independently of any specific measurement problem.
- One stage in which the measuring system is customized and applied to a specific measurement problem (measurand definition, instrument operation and data processing).

Both the whole process and the two stages individually can be performed in iterative way: through subsequent refinements. An even broader picture would lead to include in the framework at least two more components:

- the definition of the general property, which we have taken for granted here but which would need to be established in cases of new (as-yet undefined) properties, and which is particularly common in social measurement where is usually considered an inherent part of the measurement process as such;
- the specification of the objectives for which the measuring system is firstly developed and then applied, and then, complementarily, the validation of the obtained results.

Some more considerations.

First, the structural description presented here applies to nominal properties, i.e., the algebraically weakest type, but can easily specialized to ratio quantities, e.g., by transforming the lookup table exploited in the procedure \([P]\) in a calibration function defined in usual analytical form.

Second, this structural description clearly shows why instrument calibration and measurement are complementary but distinct processes: the former operates on an object – a measurement standard – whose property under consideration is assumed to be known, whereas the latter operates on an object – the object under measurement – whose property under consideration is assumed to be unknown and in fact looked for by means of measurement itself. Calibration is aimed at constructing a calibration table / function, whereas measurement at using it in its inverse form.

Third, this structural description applies independently of the nature of involved properties, either physical or non-physical, and does not impose the requirements of strongly defined measurement. Hence it could be assumed as characterizing the encompassing concept that Finkelstein termed “widely defined measurement” \([1]\): as such, it appears an appropriate ground for an operative concept of measurement that is shareable across multiple disciplinary domains.

REFERENCES