Measurability

Luca Mari
Università Cattaneo – LIUC – Italy
for "Measurement in Economics"

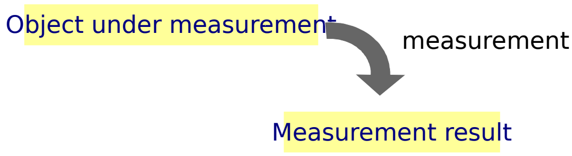
Abstract

Words have nothing magic in them: there are no "true words" for things, nor "true meanings" for words, and discussing about definitions is usually not so important. Measurement assumed a crucial role in physical sciences and technologies not when the Greeks stated that "man is the measure of all things", but when the experimental method adopted it as a basic method to acquire reliable information on empirical phenomena / objects. What is the source of this reliability? Can this reliability be assured for information related to non-physical properties? Can non-physical properties be measured, and how? This paper is devoted to explore these issues.

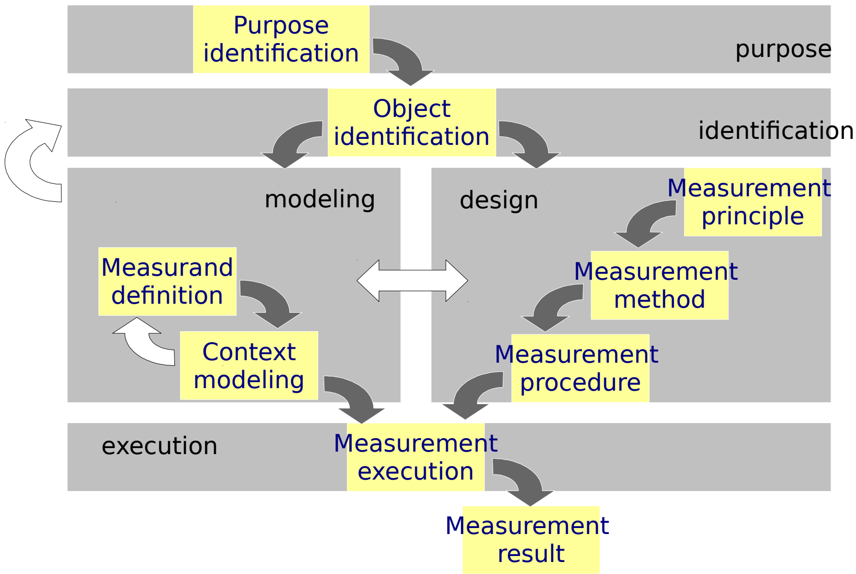
Table of Contents

1. Introduction

Measurement is an experimental and formal process aimed at obtaining and expressing descriptive information about the property of an object (phenomenon, body, substance, ...). Because of its long history and its so diverse fields of application (see at this regards [Morgan 2007]), the concept of measurement is multiform and sometimes even controversial. Indeed, while the black box model:



would interpret it as a “basic”, and actually trivial, operation, measurement can become a complex and theory-laden process, as sketched in the following diagram [Carbone et al. 2006]:



For this reason I will discuss here about measurability according to a bottom-up strategy: starting from what I suggest to be the simplest form of measurement (so simple that in fact someone could even consider it not measurement at all) I will add, step by step, some of the elements leading to a more complete framework for understanding the concept. The basic theses of the paper are:

- measurability is a specific case of evaluability;
- the measurability of a property *conceptually* depends on the current state of the knowledge of the property, and therefore it is not an “intrinsic characteristic” of the property;
- the measurability of a property *operatively* depends on the availability of experimental conditions, and therefore it cannot be derived solely from formal requirements;
- the measurement of a property is an evaluation process aimed at producing intersubjective and objective information; accordingly, measurement is a fuzzy subcategory of evaluation: the more an evaluation is / becomes intersubjective and objective, the more is / becomes a measurement.

Although somehow discussed in the following pages, I will assume here as primitive the concepts of (1) property, (2) relation among objects and properties (variously expressed as “property of an object”, “object having a property”, “object exhibiting a property”, “property applicable to an object” ...), and (3) description related to a property. Objects under measurement are considered as empirical entities, and not purely linguistic / symbolic ones, and as such the interaction with them requires an experimental process, not a purely formal one: many of the peculiar features of measurement derive from its role of *bridge between the empirical realm*, to which the object under measurement belongs, *and the linguistic/symbolic realm*, to which the measurement result belongs. I do not think that words have something magic in them: there are no "true" words for things, and discussing about definitions is usually not so important. Accordingly, I surely admit that the same term *measurement* can be adopted in different fields with (more or less) different meanings, and I do not think that the identification of a unified concept of measurement is necessarily a well-grounded aim for the advancement of science. On the other hand, a basic, historical, asymmetry can be hardly negated:

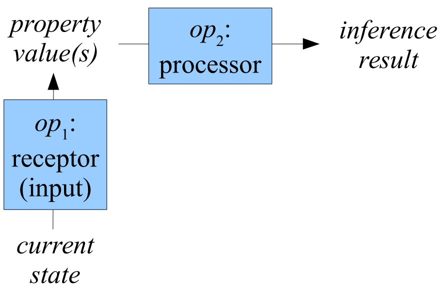
- measurement assumed a crucial role in Physics not when the Greeks stated that "man is the measure of all things", nor when they decided to call "measure" the ratio of a geometrical entity to a unit, but when the experimental method adopted it as a basic method to acquire reliable information on empirical phenomena / objects;
- for many centuries measurement has been exclusively adopted in the evaluation of physical properties, and it is only after its impressively effective results in this evaluation that it has become a coveted target also in social sciences.

As a consequence, I will further assume that:

- a *structural* analysis of the measuring process for physical properties should be able to highlight the characteristics which guarantee the intersubjectivity and the objectivity of the information it produces;
- as far as the analysis is maintained at a purely structural level, *its results should be re-interpretable for non-physical properties*.

1.1. Measurement as tool for inference

As any production process, measurement can be characterized by its aims. I suggest that measurement is primarily *a tool for inference*, whose structure can be sketched as:



where:

- op_1 is a *sensing* operation, by which some information on the current state of a system, in the form of values of one or more of its properties, is acquired;
- op_2 is a *processing* operation, by which some conclusions is inferentially drawn from such values.

The following examples of this process introduce in an evolutionary way some of the topics which will be further addressed in this paper.

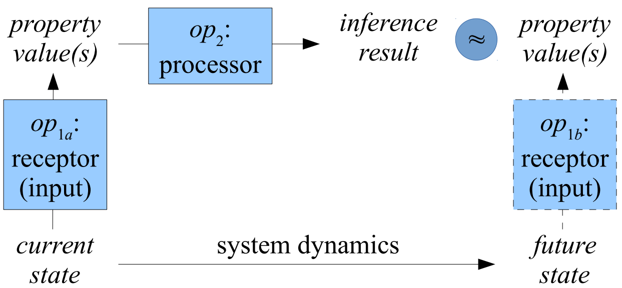
Case 1 – Two subsystems, x_1 and x_2, are identified, and the same property p is evaluated on them (op_1), thus obtaining the values $p(x_1)$ and $p(x_2)$. From the comparison of these values the inference can be drawn (op_2) whether x_1 and x_2 are mutually substitutable as far as the given property is concerned. As the resolution of the evaluation process increases (e.g., typically by increasing the number of the significant digits by which the values $p(x_i)$ are expressed), the inference result is enhanced in its quality.

Case 2 – The property p leads to a meaningful comparison in terms not only of substitutability or non-substitutability, as in the *Case 1*, but also of ordering. Hence, from $p(x_1)<p(x_2)$ the inference can be drawn that x_1 is "empirically less" than x_2 with respect to p. As the structure of the meaningful comparisons increases (e.g., typically by identifying a p-related metric among subsystems), the inference result is enhanced in its quality.

Case 3 – The values of $n\geq 2$ properties p_i of the system x are constrained by a mathematical expression, let us assume of the form $p_n=f(p_1, ..., p_{n-1})$. If the properties $p_1, ..., p_{n-1}$ are evaluated on x, then the inference can be drawn that the value of the property p_n of x is $f(p_1(x), ..., p_{n-1}(x))$. Moreover, if time variability of the properties p_i is taken into account, $p_i(x)=p_i(x(t))$, and the mathematical expression has the differential form:

$$\frac{dp_n}{dt}=f(p_1, ..., p_{n-1})$$

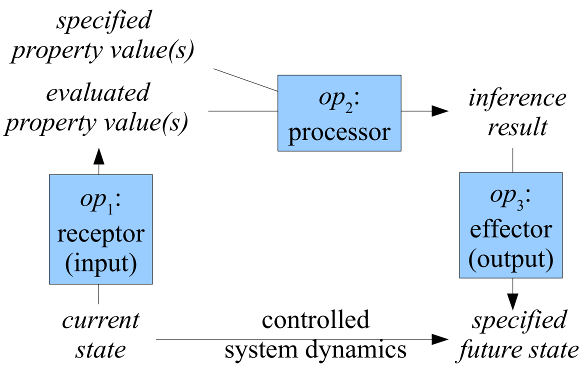
sometimes called *canonical representation* for a dynamic system (it can be noted that several physical laws have this form, possibly as systems of such first-order differential equations), then the inference becomes a prediction. The diagram:



shows the basic validation criterion in this case: the values $p_n(x(t_{future}))$ obtained in $t_{current}$ as inference result ($op_{1a}+op_2$) and by directly evaluating p_n in t_{future} (system dynamics + op_{1b}) must be compatible with each other.

Furthermore, one or more property values could be specified as nominal (or target) values instead of being evaluated: in the *Case 1*, for example, this would lead to compare the values $p(x)$ and $p(nominal)$ to establish whether x is in conformance with the given specifications with respect to p.

The inference result can be then interpreted as a decision on how to operate on $x(t_{current})$ so to obtain the specified state $x(t_{future})$:



where therefore op_3 is an *actuation* operation.

In decision-making terms, the empirical outcome of having the current state $x(t_{current})$ transformed to a different state $x(t_{future})$ is achieved by acquiring some information on the current state, then processing this information together with the specifications which express the target values, and finally operatively carrying out the decision. This structure shows a general, pragmatic, constraint put on op_1 and op_2, as expressed in terms of the commutativity of the previous diagram: the empirical transformation $x(t_{current}) \rightarrow x(t_{future})$ and the composition $op_1 + op_2 + op_3$ must be able to produce the same results. Indeed, since the operation op_3 requires some empirical transformations to be performed, the good quality of the result of $op_1 + op_2$ is not sufficient to guarantee the good quality of the final outcome. On the other hand, a low quality empirical outcome must be expected from a low quality result of $op_1 + op_2$, a principle sometimes dubbed GIGO, “garbage in, garbage out” (this does not imply, of course, a related necessary condition, given the evidence that sometimes the wanted empirical outcome is obtained even from wrong decisions: since I am interested in arguing here about measurement, and not good luck, intuition, role of individual experience, ..., in decision, I will not deal with this kind of situations here).

In the jargon of the physical sciences and technologies, op_1 can be performed as a *direct measurement*, whereas a direct measurement followed by a op_2 inference is called a *derived* (or also *indirect*) *measurement*. In this sense, a data processing operation is recognized to be a possible component of measurement, provided that at least some of its inputs come from a direct measurement (and not only from specifications, guesses, ...).

2. A basic model of measurement

Not every object has every property. Given a property p, the *domain* of p, $D(p)$, is the set of objects $\{x_i\}$ having the property p, so that $x \in D(p)$ asserts that the object x has the property p. For example, if p is the property “length” then physical rigid objects usually belong to $D(p)$, in the sense that they have a length, but social objects such as organizations do not, since they do not have a length. For a given property p and a given object x in $D(p)$, the descriptive information on p of x is denoted as $v=p(x)$ and it is called the *value v* of p of x, as in the syntagm “the value of the length of this table”, expanded but synonymous form of “the length of this table”. Values of properties can be simple entities as booleans, as in the case of the property “1 m length”, or they can be, for example, vectors of numbers, as for the property “RGB color” by which each color is associated with a triple of positive numbers. The set $V=\{v_i\}$ of the possible values for *p must contain at least two elements*, so that the assertion $p(x)=v_i$ conveys a non-null quantity of information, provided that the a priori probability of the assertions $p(x)=v_j$, $i\neq j$, is positive, so that $p(x)=v_i$ reduces the (objective or subjective) current state of uncertainty on the property value [(1)].

Properties can be thus interpreted *as* (conceptual and operative) *methods to associate values to objects*. Accordingly, the diagram:

$$\boxed{x} \xrightarrow{p} v$$

graphically expresses the fact that $p(x)=v$.

In this paper the following terminology will be adopted (see also [ISO 1993]):

- measurement is a *process* aimed at assigning a value to a property of an object;
- the measured property is called a *measurand*; measurands can be both physical and non-physical properties; they can be as simple as the length of a rigid rod or as complex as the reliability of an industrial plant;
- the value $p(x)$ obtained by measuring a measurand p of an object x expresses the result of this measurement: therefore a *measurement result* is a descriptive information entity on a property p of an object x.

The basic concept for operatively characterizing properties is *mutual substitutability*: distinct objects can be recognized as mutually substitutable in attaining a purpose. For example, objects which are different in shape, color, ... can be recognized as substitutable with each other as far as the purpose of filling a given round hole is considered. This recognition requires an experimental comparison to be performed among candidate objects, aimed at assessing their mutual substitutability. Since it does not involve any information handling, such a process is integrally empirical, and as such it can be considered as a primitive operation. *Properties can be operationally*

1 This standpoint has been formalized in terms of a concept of *quantity of information* [Shannon 1948]. The quantity of information $I(v)$ conveyed by an entity v depends inversely on the probability $PR(v)$ assigned to v: as $PR(v)$ decreases, $I(v)$ increases. From a subjective standpoint, $I(v)$ expresses the “degree of surprise” generated by the entity v. The boundary conditions, $PR(v)=1$ (logical certainty) and $P(v)=0$ (logical impossibility), correspond respectively to null and infinite quantity of information conveyed by v. Hence, an entity v brings a non-null quantity of information only if V contains at least a second element v', such that $I(v')>0$. The formal definition, $I(v)=-\log_2(PR(v))$ bit, only adds a few details to this conceptualization.

interpreted in terms of this concept of mutual substitutability: if two objects, x_1 and x_2, are recognized as mutually substitutable, then there exists a property p such that both x_1 and x_2 belong to $D(p)$, and their mutual substitutability is the empirical counterpart of $p(x_1)=p(x_2)$ (this position endorses a generalized version of operationalism, whose original characterization, "the concept is synonymous with a corresponding set of operations" [Bridgman 1927], has been acknowledged as too narrow; indeed, nothing prevents here that the same property is evaluated by different operations). As a consequence, for a given property p, an experimental *comparison process* $c_p(x_1,x_2)$ can be available:

$$\boxed{x_1} \xleftrightarrow{\;c_p\;} \boxed{x_2}$$

such that two objects x_1 and x_2 in $D(p)$ can be compared relatively to p. The process c_p is formalized as a relation, so that $c_p(x_1,x_2)=1$ means that x_1 and x_2 are recognized in the comparison substitutable with each other as far as p is concerned, the opposite case being $c_p(x_1,x_2)=0$, where 1 and 0 correspond thus to the boolean values 'true' and 'false' respectively:

$$\textit{either}\quad \boxed{x_1} \xleftrightarrow{\;c_p=1\;} \boxed{x_2} \qquad \textit{or}\quad \boxed{x_1} \xleftrightarrow{\;c_p=0\;} \boxed{x_2}$$

In the simplest case the result of this comparison is formalized as an equivalence relation: together with the immaterial condition of reflexivity, $c_p(x,x)=1$, and the usually non critical condition of symmetry, $c_p(x_1,x_2)=1$ if and only if $c_p(x_2,x_1)=1$, the relation c_p is assumed to be transitive, if $c_p(x_1,x_2)=1$ and $c_p(x_2,x_3)=1$ then $c_p(x_1,x_3)=1$. In more complex situations, both the requirements of symmetry and transitivity can be removed, and c_p can even be formalized as a non-classical, multi-valued / fuzzy, relation. In particular, in the case c_p is assumed as a non symmetric relation the following representation will be adopted:

$$\boxed{x_1} < \!\!\xrightarrow{\;c_p\;} \boxed{x_2}$$

2.1. Conditions for measurement

Measurement is recognized to be a peculiarly effective operation for obtaining descriptive information on objects. In the course of history the *reasons of this effectiveness* have been looked for in both *ontological* characteristics of the object and *formal* characteristics of the process:

- measurement has been traditionally founded on the hypothesis that properties have a "true value", i.e., a value inherently existing in the object, which measurement has the ability to determine or at least to approximate when errors are experimentally superposed to it;
- more recently, measurement has been characterized as a process by which one or more experimentally observed relations among objects are represented by formal relations defined

among property values.

With respect to other processes having comparable goals, measurement claims the ability of producing information that is reliably *intersubjective and objective* [Mari 2003]:

- *intersubjectivity* of measurement implies that its results can be interpreted in the same way by different subjects, who from the same measurement result are able to infer the same information on the measurand; this concept of intersubjectivity corresponds formally to non-ambiguity and organizationally to harmonization;
- *objectivity* of measurement implies that its results convey information only on the object under measurement and the measurand, and not on the surrounding environment, which also includes the subject who is measuring, nor on any other property of the object.

This claim of intersubjectivity and objectivity is founded on the *structural characteristics* of the measurement process: it is precisely the fact that measurement can be characterized in a purely structural way, therefore not considering any requirement on the usage of physical devices, that leaves the issue of measurability open to both physical and non-physical properties. Accordingly, measurement is *ontologically-agnostic*: in particular, it does not require measurands to have a “true value”, however this concept is defined, although it does not prevent and is usually compatible with this hypothesis. The empirical content of intersubjectivity and objectivity cannot be guaranteed to measurement by formal constraints, with the consequence that *any purely formal characterization of measurement cannot be complete*. This applies to both the classical definition of measurement as ratio to a unit and the current representational definition of measurement as scale homomorphism [Michell 2007]. Any model of measurement should be able to describe the *structure of the measurement process* as a means to obtain and express intersubjective and objective information on measurands.

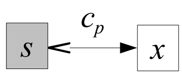
2.2. Structure of the measurement process

In its simplest structure, the measurement process of a measurand p of an object x in $D(p)$ can be described as follows:

1. *preliminary stage* (“reference construction”): an object s (“reference object”) is chosen in $D(p)$ such that:
    - the value $v_1=p(s)$ (“reference value”) is assumed to be known, possibly because conventionally chosen; together with v_1, the value set V for p contains a second value, v_0;

$$\boxed{s} \xrightarrow{p} v_1$$

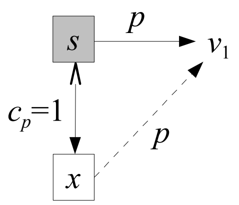
    - an experimental comparison process is available, by which s can be compared to other objects x in $D(p)$, so that the value $c_p(s,x)$ can be determined;



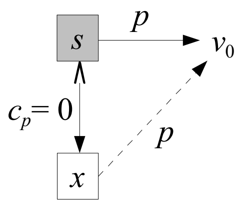
2. *determination (experimental) stage*: the object under measurement x is compared to the reference object s, and the value $c_p(s,x)$, either 1 or 0, is experimentally determined;
3. *assignment (symbolic) stage*: the value $p(x)$ is assigned according to the rule: if $c_p(s,x)=1$ then $p(x)=v_1$, else $p(x)=v_0$ [Mari 1997],

where thus a complete measurement process requires a calibration (first stage) and a measurement (second and third stage). Therefore:

either



or



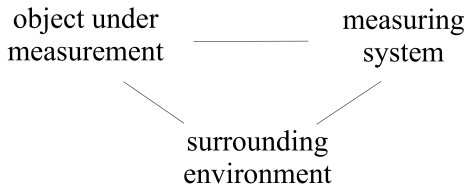
The result for this measurement process can be thus expressed as "$p(x)=v$ in reference to s by means of c_p". Whenever distinct comparisons regularly produce the same value, i.e., $c'_p(s,x)=c''_p(s,x)$ even if $c'_p \neq c''_p$, then the last specification can be removed and measurement results are expressed more customarily as "$p(x)=v$ in reference to s".

The previous diagrams assume the simplified situation in which calibration and measurement are performed synchronously, $t_{cal} = t_{meas}$: this is seldom the case. More generally, the reference object s should be then identified in its state, $s=s(t)$, that can change during time, i.e., $s(t_{cal}) \neq s(t_{meas})$ and therefore $c_p(s(t_{cal}),s(t_{meas}))=0$. This highlights the *inferential structure of measurement*:

$$\text{if} \quad \underbrace{v=p(s(t_{cal}))}_{\text{calibration}} \quad \text{and} \quad \underbrace{c_p(s(t_{meas}),x)=1}_{\text{comparison determination}} \quad \text{and} \quad \underbrace{c_p(s(t_{cal}),s(t_{meas}))=1}_{\text{reference stability}} \quad \text{then} \quad \underbrace{p(x)=v}_{\text{assignment}}$$

and explains why the basic requirement on reference objects is their stability.

The experimental and logical components required to perform a measurement process of a measurand p of an object x (at least: a reference object s associated to a reference value and a physical or logical device to perform the comparison process leading to determine the value $c_p(s,x)$) together constitute a *measuring system*. Both the object under measurement and the measuring system are embedded in an environment, whose presence generally influences the interaction of the former two elements. Hence, a measurement process involves three mutually interacting entities: an object under measurement, a measuring system, and a surrounding environment.



Accordingly, measurement can be thought of as *a process aimed at formally expressing the result of the experimental comparison of an object to a reference relatively to a property, performed by the suitable usage of a measuring system which interacts with the object in its environment*.

The information obtained by a measurement process is:

– the more *intersubjective* the more the reference object s is stable and widely available, so that the information obtained in the comparison can be transferred over the time and the space, and multiple subjects can perform the comparison and obtain the same results;

– the more *objective* the more the comparison c_p produces a value $c_p(s,x)$ depending only on the stated entities (the property p, the reference object s, and the measured object x), and not on any other entity of the surrounding environment.

Intersubjectivity and objectivity are thus interpreted as varying in a gradual, instead of sharp, way. By assuming the compliance to this structure as a requirement for a process to be considered a measurement, a concept of *quality of measurement* derives, such that different measurement processes can lead to results of different quality, in terms of their intersubjectivity and objectivity. It is indeed the quest for this quality that justifies an important part of the research and development done in Measurement Science and Technology.

2.3. Pragmatics of measurement

Among the various *pragmatic reasons* for which measurement is performed, two of them deserve specific attention for their structural implications: measurement as a first stage of an inferential process, and measurement as a means for determining the mutual substitutability of objects.

Measurement can be performed as a *first stage of an inferential process*: a value $p(x)$ obtained by means of a process structured as presented above can be adopted as a premise in an inference of the form "if $p(x)=v_1$ then $q(x)=w_1$, else $q(x)=w_0$", where q is a property such that $x \in D(q)$ and $W = \{w_0,w_1\}$ is the set of its possible values. In a more general form, the premise of the inference includes the conjunction of two or more expressions "$p_i(x)=v_j$", each of them related to a distinct property: "if $p_1(x)$=... and $p_2(x)$=... and ... then ...". Physical laws are examples of this inferential structure, stating a mutual connection, and therefore a regularity, among the involved properties. The whole process of measurement of the measurands p_i together with the application of the inference rule is called *derived* (or *indirect*) *measurement*. In the case the property evaluated by the inference can be in its turn independently measured relatively to a given reference object, the comparison between the results obtained in the two situations for the same object x can be abductively adopted to validate both the processes.

Measurement can also be performed as a *means for determining the mutual substitutability of distinct objects*: in the case $c_p(s,x_1)=c_p(s,x_2)$, and therefore $p(x_1)=p(x_2)$, where x_1 and x_2 are distinct objects in $D(p)$, such objects can be inferentially assumed to be substitutable with each other with respect to p, i.e., $c_p(x_1,x_2)$. Such an inference requires a peculiar form of transitivity of the relation c_p, i.e., $c_p(s,x_1)$ and $c_p(s,x_2)$ implies $c_p(x_1,x_2)$, which becomes a transitivity in the case c_p is symmetric. The fundamental role of this property is witnessed by the first axiom in Book I of the Euclid's Elements: "Things which equal the same thing are equal to one another".

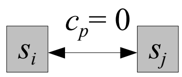
3. Extensions to the basic model

We are now ready to introduce some extensions to the simple structure presented above, with the aim of characterizing the measurement process with more details and realism.

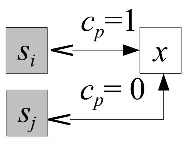
3.1. Reference as a set

The information, that relatively to the measurand the measured object is either equivalent or not equivalent to the chosen reference object, can be sometimes refined, i.e., *quantitatively* increased. A whole set of reference objects, $S=\{s_i\}$, $i=1,...,n$, called a *reference set*, can be chosen such that:

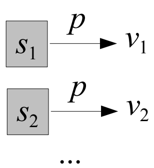
- the reference objects can be compared to each other with respect to the measurand, and any two distinct objects in S are not equivalent to each other, $c_p(s_i,s_j)=0$ if $i\neq j$, i.e., the objects in S are mutually exclusive with respect to c_p;



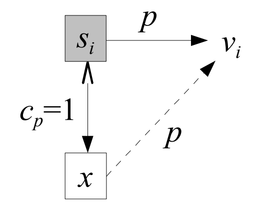
- each reference object in S can be compared to the object under measurement x, and $\exists! s_i$ such that $c_p(s_i,x)=1$, being $c_p(s_j,x)=0$ for all other $s_j\neq s_i$, i.e., the objects in S are exhaustive with respect to c_p;



- to each reference object $s_i \in S$ a value $v_i=p(s_i)$ is associated; the value set V for p is then assumed to be $\{v_i\}$, $i=1,...,n$.



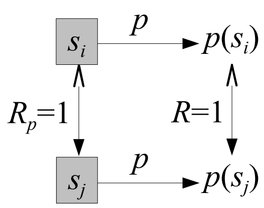
Hence, the measurand value for x is assigned so that if $c_p(s_i,x)=1$ then $p(x)=v_i$, and the measurement result is therefore expressed as "$p(x)=v_i$ in reference to S".



3.2. Reference as a scale

The information, that relatively to the measurand the measured object is equivalent to an element of the chosen reference set, can be sometimes *qualitatively* enhanced. The elements of the reference set S can be compared to each other with respect to an experimental, measurand-related, relation R_p. Assuming R_p to be binary for the sake of notation simplicity, such a relation is such that:

- for each couple (s_i,s_j) of elements in S the fact that either $R_p(s_i,s_j)=1$ or $R_p(s_i,s_j)=0$ can be determined, i.e., R_p is complete on $S\times S$;
- a relation R among the elements of the value set V is present in correspondence to R_p, such that, for each couple (s_i,s_j) of elements in S, $R_p(s_i,s_j)=1$ implies $R(p(s_i),p(s_j))=1$, i.e., the experimental information obtained in the comparison process R_p is preserved by R when expressed in terms of property values;



- the relation R_p is transferred by c_p to the objects under measurement $x\in D(p)$, so that if $R_p(s_i,s_j)=1$ and $c_p(s_i,x_i)=1$ and $c_p(s_j,x_j)=1$ then $R_p(x_i,x_j)=1$ and therefore $R(p(x_i),p(x_j))=1$.

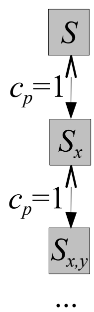
A common relation for which such conditions hold is the experimental *ordering* $<_p$, such that if $s_i<_p s_j$ (the usual infix notation for $<_p(s_i,s_j)=1$) then $p(s_i)<p(s_j)$: that is why a reference set S equipped with a relation R_p is called a *reference scale*. In this case measurement results are expressed as "$p(x)=v$ in reference to $<S,R_p>$", where the couple $<S,R_p>$ is called a relational system (in more general terms, a set $\{R_p\}$ of relations could be defined on S, so that the relational system is the couple $<S,\{R_p\}>$; this extension, immaterial for the present discussion, is the main topic of the already mentioned representational definition of measurement [Michell 2007]).

Considered as a first stage of an inferential process, measurement based on a reference scale leads to values that can be adopted as premises in inferences such as "if $p(x_i)=$... and $p(x_j)=$... and $R(p(x_i),p(x_j))=1$ then $R_p(x_i,x_j)=1$", which thus *exploit the structure* induced by R_p on $D(p)$.

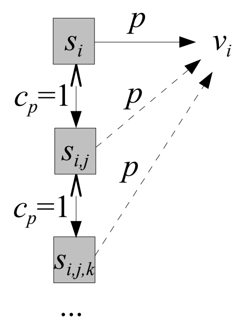
Since an n-ary operation is a specific (n+1)-ary relation, R_p can be sometimes expressed as an *operation* Op_p. Assuming Op_p to be binary for the sake of notation simplicity, R_p is ternary and $R_p(s_i,s_j,s_k)=1$ if and only if $Op_p(s_i,s_j)=s_k$. Therefore a binary operation Op on the value set V corresponds to Op_p, such that if $Op_p(s_i,s_j)=s_k$ then $Op(p(s_i),p(s_j))=p(s_k)$. Operatively important is the situation in which the binary operation Op has the properties of a sum among values in V, and a procedure for the experimental replication of the reference objects in S is available, i.e, the reflexive property $c_p(s_i,s_i)=1$ assumes the operative meaning that the reference object s_i has a clone. In this case if $Op_p(s_i,s_i)=s_j$ then $p(s_i)+p(s_i)=p(s_j)$, and therefore $p(s_j)=2p(s_i)$. A reference object s_1 can be then chosen so that $p(s_1)=1$, with the role of *scale unit* by which the reference objects can be operatively generated: $s_2=Op_p(s_1,s_1)$, $s_3=Op_p(s_2,s_1)=Op_p(Op_p(s_1,s_1),s_1)$, ... and $p(s_n)=np(s_1)$. Hence, in this case measurement results can be expressed as "$p(x)=n$ in reference to s_1", being s_1 the chosen scale unit.

3.3. Traceability

The goal of enhancing the intersubjectivity of measurement can be obtained by increasing the number of experimental situations, distinct in space and / or in time, in which the *same* reference S (either a set or a relational system, possibly equipped with a unit) is adopted. This requires S to be available to perform the comparisons $c_p(s_i,x)$ which constitute the experimental component of measurement. This problem is commonly dealt with by experimentally generating some replicas S_x of S, and then iteratively generating some replicas $S_{x,y}$ of the replicas S_x until required, and finally disseminating these replicas to make them widely available (according to this notation, $S_{x,y,z}$ is therefore the z-th replica of the y-th replica of the x-th replica of S). The whole system of a reference S and its replicas is therefore based on the assumption that $c_p(S,S_x)=1$ and that, iteratively, $c_p(S_x,S_{x,y})=1$, having suitably extended the relation c_p to sets and relational systems. Hence, an *unbroken chain* $c_p(S,S_x)=1$, $c_p(S_x,S_{x,y})=1$, $c_p(S_{x,y},S_{x,y,z})=1$, ... makes the last term traceable to S, which thus has the role of primary reference for all the elements of the chain:



such that for example:



Under this assumption of *traceability*, measurement results are still expressed relatively to the primary reference S, even if this relation is only indirect, being based on the transitivity of the relation c_p. The basic characteristic of any given replica $S_{x,y}$ of a reference S_x is the guarantee that $c_p(S_x,S_{x,y})=1$. This is an experimental, not a formal, fact, which must be ascertained by operatively comparing the two references, and modifying the state of the replica $S_{x,y}$ in the case $c_p(S_x,S_{x,y})=0$, an operation called *reference calibration* (on the concept of calibration see also [Boumans 2007]).

As a traceability system grows in number of the disseminated replicas, its primary reference becomes more and more intersubjectively important in the measurement of the property under consideration: therefore, the quality of measurement is not only a characteristic of the specific process under consideration but also implies some systemic components. A reference in a socially widespread traceability system is called a *measurement standard* (or simply "standard"). The primary reference in a traceability system of measurement standards is called "primary standard", and the references used to operatively perform measurements (instead of disseminating the primary standard) are called "working standards". Hence, in a traceability chain the first and the last element are the primary standard and a working standard respectively. The adoption of a traceability system modifies the structure of a measurement process as follows:

1. *preliminary stage*, in which the adopted standard S is calibrated, i.e., its traceability to a given standard is established, to assign a value $p(s_i)$ to each reference object s_i in S;
2. *determination stage*, in which the reference object s in S is identified such that $c_p(s,x)=1$, being x the object under measurement;
3. *assignment stage*, in which the value $p(x)=p(s)$ is assigned.

3.4. Asynchronous comparison by means of a calibrated sensor

Despite of the dissemination of standards by a traceability system, in some routine measurements the object under measurement could not be directly compared to a standard S to determine $c_p(s_i,x)$, because of the local unavailability of a standard and / or even the unavailability of an experimental comparison process c_p which is synchronously applicable to s_i and the object under measurement x. In these situations, measurement results can be sometimes obtained by means of an *asynchronous comparison* between the object under measurement and an available standard, through the mediation of a *measuring transducer*, usually called a sensor, i.e., a device d such that:

- d has an "output property" q which can be measured;
- d is able to interact with the objects in $D(p)$, by modifying the value of its output property in function of the value of p, which thus operates in this case as an "input property".

Because of this behavior, d is interpreted as a device transducing the measurand p to the output property q, so that the structure of the measurement process is modified as follows:

1. *preliminary stage* ("sensor calibration"): the sensor d is systematically put in interaction with an available standard S whose objects $s \in D(p)$, and for each $s \in S$ the resulting value $q(s)$ is measured; since the values $p(s)$ are assumed to be known, the set of the couples $<p(s), q(s)>$, i.e., <measurand value, corresponding output property value>, is recorded ("calibration data"), possibly in the form of a "calibration function" c, $c(p(s))=q(s)$; such a function associates a value for the sensor output property to each measurand value obtained for a working standard (note that this characterization is in accordance with the definition of calibration given by the International Vocabulary of Basic and General Terms in Metrology (VIM) [ISO 1993]: "set of operations that establish, under specified conditions, the relationship between values of quantities indicated by a measuring instrument or measuring system, or values represented by a material measure or a reference material, and the corresponding values realized by standards");
2. *transduction stage*: d is put in interaction with the object under measurement x, and the corresponding value $q(x)$ is obtained, called "indication", or "instrumental reading", i.e., the value of the sensor output property for the object under measurement;
3. *assignment stage*: the value $p(x)$ is assigned according to the rule: the couple $<p(s), q(s)>$ is found in the calibration data such that $q(x)=q(s)$, and then $p(x)=p(s)$; this assignment assumes the calibration function c to be invertible, so that from the indication $q(x)$ the measurand value $p(x)$ is assigned such that $p(x)=c^{-1}(q(x))$.

This justifies the interpretation according to which *measurement and calibration are inverse operations*. Furthermore, a calibrated sensor embeds the information on the standard against which it has been calibrated. This implies that calibrated sensors can functionally operate as standards of a traceability system.

4. Relations with the representational point of view

I have already mentioned the current representational definition of measurement as scale homomorphism. Given the status of "orthodox" measurement theory of this point of view, it is worth highlighting the elements for which the concept of measurement presented here differs from the representational one. To this goal, I suggest that measurement, as any complex operation, *can be described according to multiple levels of abstraction*. Beginning from a general description, more and more specific characterizations can be obtained such that:

- each level specializes the previous one, being included in it as a special case;
- each level highlights some features of the operation that were ignored at the previous level.

I conceive of four "levels of description" for measurement, as follows:

- Level A: measurement as generic evaluation;
- Level B: measurement as homomorphic evaluation;
- Level C: measurement as homomorphic evaluation resulting from an experimental comparison to a reference;
- Level D: measurement as empirical operation.

In its most abstract interpretation, let us call it *Level A description*, measurement is simply meant to be a generic evaluation, aimed at assigning a symbol v chosen from a given set V to any candidate object x of a set X and therefore formalized as a function $p : X \rightarrow V$. Such a function admits two complementary interpretations, as it can be thought of as representing:

- the property of the objects in X whose evaluation is expressed by means of the symbols of V;
- the operation by which the objects in X are mapped to the symbols in V.

Any function p induces an equivalence relation $\approx_p$ on its domain, such that $x_i \approx_p x_j$ if and only if $p(x_i)=p(x_j)$, i.e., two objects are equivalent if and only if they are associated to the same value by p: the subset of the objects x_i which are in this sense equivalent is an equivalence class, and therefore an element of a partition of X, usually denoted by $X/\approx_p$. As a consequence, any function p can be decomposed into:

- a "partition function" $\pi_p : X \rightarrow X/\approx_p$, which maps any object to its equivalence class;
- a "labeling function" $\lambda_p : X/\approx_p \rightarrow V$, which maps any equivalence class to a value, and such that $p(x) = \lambda_p(\pi_p(x))$.

$$\begin{tikzcd} X \arrow[rr, dashed, "p"] \arrow[dr, "\pi_p"'] & & V \\ & X/\approx_p \arrow[ur, "\lambda_p"'] & \end{tikzcd}$$

This decomposition formally justifies the initial assertion on the role of measurement of bridge between the empirical realm and the linguistic/symbolic realm: the measurement of a property p of an object x corresponds to the empirical determination of the $\approx_p$-equivalence class which x belongs to followed by the symbolic assignment of a value to this class. I see this as the main merit of the Level A description, which on the other hand is unable to specify any constraint on the evaluation (note that λ_p is 1-1 by definition), thus leading to a far too generic description of measurement.

The available knowledge on the property p could guarantee that among the objects in X one or more relations related to p can be observed together with the $\approx_p$-equivalence. For example, objects x_i and

x_j which are not $\approx_p$-equivalent to each other could satisfy an order relation $<_p$ such that as far as p is concerned x_i is not only distinguishable from x_j, but also "empirically less" than it. In these cases the labeling function λ_p must be constrained, so to preserve the available structural information and to allow inferring that $x_i <_p x_j$ from $p(x_i)<p(x_j)$, as from $p(x_i)=p(x_j)$ the conclusion that $x_i \approx_p x_j$ can be drawn. To satisfy this further condition, p is formalized as a homomorphism: this *Level B description*, which clearly specializes the Level A description, emphasizes indeed the constraints that a consistent mapping p satisfies, as formalized by the concept of the scale type in which the property is evaluated. For example, an evaluation performed in an ordinal scale is defined but a monotonic transformation, so that if $p : X \rightarrow \{1,2,3,4,5\}$ is ordinal then the transformed mapping p' : $X \rightarrow \{10,20,30,40,50\}$ such that $p'(x)=\tau(p(x))$, where $\tau(y)=10y$, conveys exactly the same information as p. Hence, each scale type corresponds to the class of the allowable transformation functions $\tau : V \rightarrow V'$ which preserve the relations defined on X [(2)].

$$\begin{tikzcd}
X \arrow[rr, dashed, "p"] \arrow[rrr, dashed, bend left, "p'"] \arrow[dr, "\pi_p"'] & & V \arrow[r, "\tau"] & V' \\
& X/\approx_p \arrow[ur, "\lambda_p"'] \arrow[urr, "\lambda'_p"'] & &
\end{tikzcd}$$

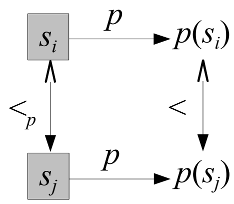
I see this link with the concept of scale type as the main merit of the Level B description, which on the other hand is unable to specify any constraint on the evaluation that guarantees its intersubjectivity and objectivity, thus leading to a description of measurement that is still too generic. The Level B description expresses the representational point of view to a theory of measurement.

The model of measurement that has been presented in the previous pages corresponds to the *Level C description*, which characterizes measurement as a homomorphic evaluation resulting from an empirical comparison to a reference. Indeed, if a reference scale is available for the property p such that, for example, an experimental order $<_p$ is defined between reference objects, then the above specified conditions on the scale require that:

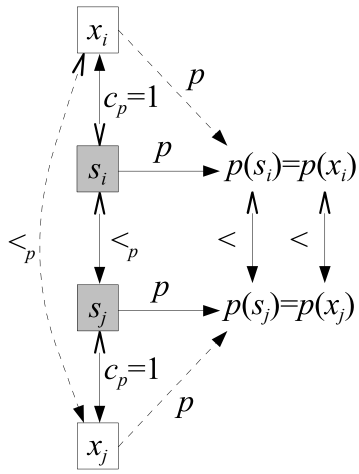
- the experimental information related to the order $<_p$ must be preserved in terms of the symbolic

2 As a corollary of this definition, it can be easily shown that the transformation functions τ are injective, i.e., map distinct arguments to distinct values. The algebraically weakest, and therefore more general, scale type is the nominal one, for which the only preserved relation is the $\approx_p$-equivalence, so that the only constraint on its transformation functions is injectivity. Each other scale type specializes the nominal one by adding further constraints to injectivity, for example monotonicity for the ordinal type and linearity for the interval type. It is precisely this common requirement of injectivity that justifies the fact that the transformation functions preserve the information acquired in the experimental interaction with the object under measurement, as expressed in the recognition of its membership to a given $\approx_p$-equivalence class.

order < defined among property values:



- the order $<_p$ is transferred by the comparison process c_p to the objects under measurement x, so that if $s_i <_p s_j$ and $c_p(s_i,x_i)=1$ and $c_p(s_j,x_j)=1$ then also $x_i <_p x_j$ and therefore $p(x_i)<p(x_j)$:



Therefore, the following chain of implications holds:

- if a reference scale is defined for p;
- and if for a given couple of reference objects s_i, s_j such that $s_i <_p s_j$ the conditions hold that $c_p(s_i,x_i)=1$ and $c_p(s_j,x_j)=1$;
- then also $x_i <_p x_j$ and therefore $p(x_i)<p(x_j)$.

This result is easily generalized to any relation R_p, and shows that the condition of homomorphism for the objects under measurement trivially follows from the condition of homomorphism for the reference objects, and therefore that the Level C description specializes the Level B one. I see the introduction of the concept of (traceable) reference and the formalization of the experimental comparison between the object under measurement and the reference as the main merit of the Level C description, which maintains an abstract connotation on the specific methods adopted to experimentally perform such a comparison.

Finally, a *Level D description* can be envisioned, which further specializes the Level C description by identifying the empirical operations performed to compare the object under measurement to the assumed reference. In the previous pages two of such methods have been introduced, namely, the synchronous direct comparison and the asynchronous comparison mediated by a measuring transducer. Descriptions of measurement methods can be found in most technical books on measurement. Also standard documents such as the International Vocabulary of Basic and General

Terms in Metrology (VIM) [ISO 1993] list them in various ways (indeed, according to the VIM, "measurement methods may be qualified in various ways such as: substitution measurement method; differential measurement method; null measurement method; direct measurement method; indirect measurement method").

If, as I am suggesting, it is the Level C description the one which specifically highlights the characteristics of measurement, the question arises whether the representational point of view (Level B) can be properly considered a theory of measurement. At this regards a serious ambiguity must be preliminarily solved, related to the status of the relations defined among the objects in X: is their observability an *experimental* requirement or just a *logical* one? That is: should the relations R_p be directly observed as the result of an experimental comparison process $R_p(x_i,x_j)$, or can their existence be inferred from the comparisons $R(p(x_i),p(x_j))$?

The Level C description only requires the experimental observability of the relations $R_p(s_i,s_j)$, that generate the reference scale, not necessarily of $R_p(x_i,x_j)$. Let us at first assume that also the Level B description allows the relations R_p to be obtained indirectly (let us call this a *weak* representational point of view). Accordingly, Level C specifies Level B, which is not a theory of measurement only because too generic: the weak representational point of view gives a *necessary* but not sufficient condition to characterize measurement. If, on the other hand, the relations among the objects in X are required to be directly observable (a *strong* representational point of view), the situation becomes more complex: a property evaluation could satisfy all the Level C requirements and at the same time the relations $R_p(x_i,x_j)$ could remain unobserved. The strong representational point of view gives neither a sufficient nor a necessary condition to characterize measurement (a further discussion on this subject can be found in [Mari 2000]) (3).

5. Quality of measurement

I have already considered that, as for any production process, measurement should be evaluated relatively to the quality of its products, i.e., measurement results. The quality of measurement results has been traditionally accounted for in terms of error, i.e., difference of the reported measurand value and the true value of the measurand itself, thus seeking an ontological solution to a pragmatic problem. I suggest that the inconclusiveness of many analyses on this topic depends on the lack of a clear distinction between the empirical realm and the linguistic/symbolic one. Indeed, the cultural tradition from which the very concept of measurement grew up, from the Pythagorean

3 In its usual interpretation, is the representational point of view *strong* or *weak*? Let us take into account a classical definition at this regards: "Measurement is the assignment of numbers to properties of objects or events in the real world by means of an objective empirical operation, in such a way as to describe them. The modern form of measurement theory is representational: numbers assigned to objects/events must represent the perceived relations between the properties of those objects/events" [Finkelstein, Leaning 1984]. This emphasis on perception seems to give a clear answer to the question...

school, to Euclid, to Galileo, to Gauss, to Kelvin, *grounded measurability on the assumption that "numbers are in the world"* (as Kepler wrote in his Letter to Michael Maestlin, 1595).
Measurement was interpreted as a process of *discovery* of entities that are already in the object under measurement, so that measurement results would be empirically determined, i.e., "extracted" from the underlying objects. Quantities would be themselves inherent characteristics of objects, the concept of true value for a quantity simply being the coherent outcome of this standpoint. On the other hand, in about the last 100 years a pivotal concept appeared, and more and more became crucial for any scientific analysis and development: *the concept of model*. The current view on symbolization can be traced back to the concept of formal system as defined by David Hilbert: theories are purely symbolic constructions, and as such they can (and should) be consistent, but they are neither true nor false since, strictly speaking, they do not talk about anything. Truth is not a property of symbols, and surely not even of empirical objects, but of models, i.e., interpretations of theories that are deemed to be true whenever they manifest themselves as empirically coherent with the given domain of observation. According to our current model-based view, numbers are *not* in the (empirical) world simply because they *cannot* be part of it. Indeed, let us compare the following two statements:

- "at the instant of the measurement the object under measurement is in a definite state";
- "at the instant of the measurement the measurand has a definite value".

While traditionally such statements would be plausibly considered as synonymous, their conceptual distinction is a fundamental fact of Measurement Science: the former expresses a usual assumption of measurement (but when some kind of ontological indeterminism is taken into account, as in some interpretations of quantum mechanics); the latter is unsustainable from an epistemological point of view and however operationally immaterial (a further discussion on this subject can be found in [Mari, Zingales 2000]).
The conceptual importance of the change implied in the adoption of the concept of model should not be underestimated. It is *a shift from ontology to epistemology*: measurement results report not directly about the state of the object under measurement, but on our knowledge about this state. Our knowledge usually aims at being coherent with the known objects ("knowledge tends to truth", as customarily said), but even a traditional standpoint, such as the one supported by the above mentioned VIM, is forced to recognize that "true values are by nature indeterminate". The experimental situation which at best approximates the concept of true value for a property is the check of the calibration of a sensor by means of a reference object. In this case, the value for the input property is assumed to be known before the process is performed, and therefore actually operates as a reference value. On the other hand, this operation is aimed at verifying the calibration of a device, not obtaining information on a measurand. Indeed, if the reference value is 2,345 m and the value 2,346 is instead experimentally obtained, then the usual conclusion is not that the

reference object has changed its state (however surely a possible case), but that the sensor must be recalibrated. Plausibly for describing this kind of peculiar situations the odd term "conventional true value" has been proposed (the concept of "conventional truth" is not easy to understand...), but it should be clear that even in these situations truth is out of scope: *reference values are not expected to be true, but only traceable*. A still conservative outcome, which is adopted more and more, is of purely lexical nature: if the reference to truth is not operational, then it can simply be removed. This has been for example the choice of the Guide to Expression of Uncertainty in Measurement (GUM) [ISO 1995], which considers the adjective "true" to be redundant and accordingly writes "the value", by dropping "true". On the other hand, the pragmatic problem of properly evaluating the quality of measurement is not solved by a linguistic choice, and therefore remains an open issue.

5.1. The truth-based view

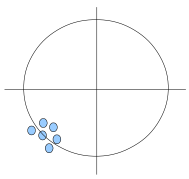
Measurement should produce information on both the measurand value and its quality, which can be interpreted in terms of reliability, certainty, accuracy, precision, ... Each of these concepts has a complex, and sometimes controversial, meaning, also because its technical acceptation is usually intertwined with its common, non-technical, usage (as a cogent example the case of the term "precision" can be considered. The VIM [ISO 1993] does not define it, and only recommends that it "should not be used for 'accuracy'", whereas it defines the repeatability as the "closeness of the agreement between the results of successive measurements of the same measurand carried out under the same conditions of measurement", called "repeatability conditions". A second fundamental standard document, also released by ISO [ISO 1998], defines the precision as "the closeness of agreement between independent test results obtained under stipulated conditions", and then notes that the repeatability is the "precision under repeatability conditions").

I do not think discussing terminology is important: words can be precious tools for knowledge, but too often discussions are only about words. *The agreement should be reached on procedures and possibly on concepts*, not necessarily on lexicon. I subscribe at this regards the position of Willard Van Orman Quine: "science, though it seeks traits of reality independent of language, can neither get on without language nor aspire to linguistic neutrality. To some degree, nevertheless, the scientist can enhance objectivity and diminish the interference of language, by the very choice of language" [Quine 1966]. Indeed, what is important for our subject is an appropriate operative modeling on the quality of measurement, not the choice of the terms adopted to describe this modeling activity and its results.

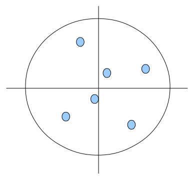
The structure of the measurement process, that in the previous pages I have introduced and then variously extended, does not include any explicit component allowing to formally derive some information about the quality of the process itself. Such a structure can be thus thought of as an "ideal" one. Two prototypical situations are then traditionally mentioned to exhibit the possible presence of “non-idealities”:

- the measurement of a property whose value is assumed to be already known (thus analogously to the check of the calibration of a sensor by means of a reference object): a difference of the obtained value from the known one can be interpreted as the effect of an error in the process, for example due to the usage of an uncalibrated sensor; this effect, which is not plausibly corrected in repeated applications of the measuring system, is traditionally called a *systematic error*;
- the measurement of a property by the repeated applications of the measuring system under the hypothesis that the state of the object under measurement does not change during the repetitions: the fact of obtaining different values in the repetitions can be interpreted as the effect of an error in the process. The superposition of several, unidentified but singularly small, causes generated by the interaction of the environment with the object under measurement and / or the measuring system is typically assumed; this effect, under the hypothesis of its statistical origin, is traditionally called a *random error*.

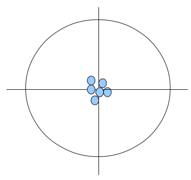
The concepts of systematic and random error and their relations are expressively exemplified by the operation of shooting at a target, as follows:



A case of systematic error with a (relatively) low random error



A case of random error with a (relatively) low systematic error



A case of (relatively) low random error and systematic error, i.e., close to “true measurement”

These pictures justify the (ideal) definition of “true value” as “the value obtained after an infinite series of measurements performed under the same conditions with an instrument not affected by systematic errors” [D’Agostini 2003]. On the other hand:

- systematic errors can be recognized as such only if a reference value for the measurand, i.e., the target point, is assumed to be known in advance; in this case, a “degree of systematic error” is evaluated by the distance (provided that a distance is algebraically defined) between the measurement result and the reference value;
- random errors can be recognized as such only if the state of the object under measurement does not change during the repetitions (in short: if the measurement is assumed to be repeatable), i.e., the target point is not a moving target; in this case, a “degree of random error” is evaluated by a dispersion index (provided that it is algebraically defined) of the set of the measurement results.

While both the hypotheses are demanding from an epistemological point of view, they are radically different in operative terms:

- a reference value for the measurand is typically not known in advance: indeed, measurement is usually aimed at obtaining available information on a measurand, and not at confirming the quality of the measuring system (which is instead the task of calibration); as a consequence, *systematic errors cannot usually be evaluated*;

- the repeatability is surely not a necessary condition for measurement, but it can be sometimes assumed as the result of the analysis of the empirical characteristics of both the measuring system and the object under measurement; as a consequence, *random errors can sometimes be evaluated*.

Apart from these epistemological issues, the traditional interpretation of quality of measurement in terms of errors is hindered by the operative *problem of formalizing these two types of error in a compatible way*, so to allow to properly combine them into a single value. None of the several solutions which have been proposed obtained a general agreement, plausibly because of their nature of ad hoc prescriptions (either "combine them by adding them linearly", or "... quadratically", or "... linearly in the case ..., and quadratically otherwise"). On the other hand, this problem has been recently dealt with in a successful way by the already mentioned GUM [ISO 1995], according to a pragmatic standpoint which is aimed at unifying the procedure and the vocabulary while admitting different interpretations of the adopted terms. In the following this standpoint will be explicitly presented, and maintained as a background reference.

5.2. The model-based view

Because of the mentioned shift from ontology to epistemology, Measurement Science emphasizes now *certainty* instead of truth. Accordingly, the quality of measurement is more and more conceptualized in terms of *uncertainty*, i.e., lack of complete certainty on the value that should be assigned to describe the object under measurement relatively to the measurand, thus acknowledging that measurement is a knowledge-based process. From a conceptual standpoint this change has some traits of a scientific revolution, in the sense of the term proposed by [Kuhn 1970]: the truth-based view and the model-based one can be thought of as competing paradigms, and some of the current problems troubling the metrological community derive from what Thomas Kuhn calls the incommensurability of such paradigms (4). On the other hand, in pragmatic terms the change from the truth-based view to the model-based one *is a domain extension*. Indeed, the uncertainty modeling does not prevent dealing with errors as a possible cause of quality degradation, but it does not force to assume that *any* quality degradation derives from errors. If measurement is not able to acquire "pure data", then it must be based on a model including the available relevant knowledge on the object under measurement, the measuring system and the measurand: this knowledge is generally required to evaluate the quality of a measurement. Indeed, several, not necessarily independent, situations of non-ideality can be recognized in the measurement process; in particular (denoting with x the object under measurement and with s the reference to which x is compared), it could happen that (5):

4 My opinion is that Measurement Science is currently living a transition phase, in which the historically dominant truth-based view is being more and more criticized and the model-based view is getting more and more support by the younger researchers. On the other hand, the truth-based view is a paradigm that benefits from a long tradition: the scientists and the technicians who spent their whole live thinking and talking in terms of true values and errors are fiercely opposing the change. An indicator of this situation is linguistic: in response to the critical analyses highlighting the lack of any empirical basis for the concept of true value, the term "conventional true value" has been introduced (the VIM [ISO 1993] defines it as "value attributed to a particular quantity and accepted, sometimes by convention, as having an uncertainty appropriate for a given purpose"). Despite its aim of extreme defense of the traditional paradigm, the very concept of "conventional truth" is so manifestly oxymoric that its adoption seems to be a cure worse than the illness. A further analysis on the current status of Measurement Science in terms of paradigms can be found in [Rossi 2006].

5 In more detailed way, the GUM mentions as "possible sources of uncertainty in a measurement": "a) incomplete definition of the measurand; b) imperfect realization of the definition of the measurand; c) non-representative sampling - the sample measured may not represent the defined measurand; d) inadequate knowledge of the effects of environmental conditions on the measurement, or imperfect measurement of environmental conditions; e) personal bias in reading analogue instruments; f) finite instrument resolution or discrimination threshold; g) inexact values of measurement standards and reference materials; h) inexact values of constants and other parameters obtained from external sources and used in the data-reduction algorithm; i) approximations and assumptions incorporated in the measurement method and procedure; j) variations in repeated observations of the measurand under apparently identical conditions".

- s is *not stable*, i.e., it changes its state during its usage, so that the value that was associated to it at the calibration time does not represent its state at the measurement time (formally: $p(s(t))=p(s(t_0))$ even if $c_p(s(t_0),s(t))=0$, a comparison that can be performed only in indirect way);
- the system used to compare x to s *not repeatable*, i.e., x and s are mutually substitutable in a given time and subsequently they result as no more substitutable even if they have not changed their state (formally: if $c_p(s(t_1),x(t_1))=1$, then $c_p(s(t_2),x(t_2))=0$ even if $c'_p(s(t_1),s(t_2))=1$ and $c'_p(x(t_1),x(t_2))=1$, where c'_p is a comparison process assumed to be more repeatable than c_p);
- the system used to compare x to the adopted reference has a *low resolution*, i.e., x is substitutable with distinct reference objects (formally: both $c_p(s_i,x)=1$ and $c_p(s_j,x)=1$ even if $s_i \neq s_j$) (this applies also to the relation between a reference and its replicas in the traceability system).

This list of situations of non-ideality does not include the item that the GUM [ISO 1995] states as the first source of uncertainty in a measurement: the incomplete definition of the measurand. Because of its relevance to the very concept of measurement uncertainty, a short analysis of *the problem of the definition of the measurand* is appropriate.

5.3. Some considerations on the definition of the measurand

As a simple example of measurand definition, let us consider the diameter of a bore, as it is presented by [Phillips et al. 2001]: “The simple definition as a diameter may be sufficient for a low accuracy application, but in a high accuracy situation imperfections from a perfectly circular workpiece may be significant”. According to a radically operational definition, this sentence is fallacious: if the measurand is defined by the operation by which it is measured, the experimental evidence of different “diameter values” obtained for different positions on the same bore would lead to the conclusion that the bore *has several diameters*, not that the diameter is incompletely defined. The concept of “having several diameters” is admittedly inconsistent with the geometrical meaning of the term, to which the mentioned “simple definition” implicitly refers: on the other hand, diameters, as geometrically defined, cannot be physical properties, for the obvious reason that in the physical world no “perfect circles” exist at all.

Radical operationalism is however seldom maintained: properties are a constitutive component of our knowledge, and we tend to assign them stable, and therefore transferable, meanings, to which any single operation only partially contributes. This dependence to a model can make a measurand definition incomplete: in the example, if the common, geometrical, concept of diameter is maintained, then “in a high accuracy situation imperfections from a perfectly circular workpiece may be significant”, and such imperfections typically lead to uncertainties in the diameter measurement. As the observation accuracy grows, the fact that diameter is not a single-valued quantity must be recognized as depending on the discrete structure of the matter, not on manufacturing imperfections anymore: in this situation, the remaining uncertainty is therefore *intrinsic to the measurand definition*. On the other hand, it is precisely the dependence on a model (instead of ontological roots) which allows the alternative option of defining measurands on ad hoc bases. For example, again [Phillips et al. 2001] notes that “some standards (...) have further defined the diameter of a bore to be the maximum inscribed diameter” (or, more plausibly, the maximum distance between points on the edge of the bore, to avoid defining the concept of diameter in terms of itself...). But this conventionality can result in arbitrariness: why not to define the diameter as the average distance between opposite points? or as the difference between the maximum and the minimum of such distances? If properties are methods to associate information entities to objects, as I have sustained above, it is possible to arbitrarily define always new properties. *Conventionality in the definition of properties avoids arbitrariness if grounded on pragmatic bases*. For example, in a system constituted of a piston and a cylinder, the internal “diameter” of the cylinder could be defined as the minimum inscribed distance and the external “diameter” of the piston as the maximum inscribed distance. Were the internal “diameter” of a cylinder c ascertained to be greater than the external “diameter” of a piston p, the passage of p through c could be inferred:

given v_1 = internal_diameter(c) and v_2 = external_diameter(p), then:

if $v_1 > v_2$ then passage(p,c)

an implication with formal analogies to, for example:

given v_1 = applied_force(x) and v_2 = mass(x), then:

if $v_3 = v_1 / v_2$ then acceleration(x) = v_3

(a rather lengthy expression of the known physical principle commonly written $F=ma$). Both cases express a law stating a relation among properties that in principle are defined independently of each other (in [Mari 1999] I have analyzed the information conveyed by this relation, by calling it “information-from-connection” and discussing its *pragmatic* nature). These relations contribute to a complex concept of definition of properties, according to which each property is partially defined by means of other properties, in a network of mutual connections expressing the available knowledge of such properties and limiting the conventionality of their definition [Mari 2005].

The network that connects the measurand to other properties guarantees the pragmatic usefulness of the measurand evaluation but, at the same time, is a further source of complexity for the definition of the measurand itself. Indeed, part of this network are the so called *influence quantities*, i.e., according to the definition of the VIM [ISO 1993], those properties of the object under measurement or the environment (thus including the measuring system) that are distinct from the measurand and nevertheless affect the measurement result. As a simple example, consider the expansion of a metallic body caused by a temperature increase: if length is the measurand, then temperature is an influence quantity. The evidence that repeated measurements on the same object produce different results because of temperature variations can be modeled according to two strategies:

- temperature is maintained as a hidden variable in the definition of the measurand, whose intrinsic uncertainty should be increased correspondingly to keep into account this under-determination;
- temperature is explicitly included in the definition of the measurand, which is then denoted for example as "length at 20 °C".

The greater specificity of this second strategy offers the potential condition of obtaining measurand values of lower uncertainty, and therefore of higher quality, but requires the measurement of two properties, length and temperature, together with a control / correction system (that can be empirical or symbolic) for dealing with the situations in which the measured temperature is different from the specified one. Moreover, in this case temperature becomes a measurand in turn, with the consequence that its value could depend on further influence quantities, for which the problem of choosing a strategy for dealing with the influence quantities is iterated. Once more, this shows the model dependence of the measurand definition.

5.4. From analysis to expression

From the previous considerations the conclusion can be drawn that measurement is not a purely empirical operation. Indeed, any measurement can be thought of as a three-stage process (see also [Mari 2005b]):

1. *acquisition*, i.e., experimental comparison of the object under measurement to a given reference;
2. *analysis*, i.e., conceptual modeling of the available information (the comparison result, together with everything is known on the measurement system: the measurand definition and realization, the instrument calibration diagram, the values of relevant influence quantities, ...);
3. *expression*, i.e., statement of the gathered information according to an agreed formalization.

The crucial role of the analysis stage is emphasized by considering it in the light of the truth-based view. Once more, were "numbers in the world", questions such as "how many digits has the (true) length of this table?" would be meaningful, while as the power of the magnifying glass increases the straight lines limiting the table become more and more blurred, and the very concept of length looses any meaning at the atomic scale. If these questions traditionally remained outside Measurement Science it is plausibly because of the impressive effectiveness that the analytical methods based on differential calculus have shown in the prevision of the system dynamics: *this led to the assumption that the "numbers in the world" are real numbers* ([6]). As a consequence, the

6 Direct consequence of this standpoint is the hypothesis that "true measurement" requires continuity, so that discreteness in measurement would always be the result of an approximation. I must confess that I am simply unable to understand the idea of numbers as empirical entities which grounds the position that in [Mari 2005] is called the "realist view": "whether a physical phenomenon is continuous or not seems to be primarily a matter of Physics, not Measurement Science. Classical examples are electrical current and energy: while before Lorenz/Millikan and Plank they were thought of as continuously varying quantities, after them their discrete nature has been discovered, with electron charge and quantum of action playing the role of ultimate discrete entities. What is the realist interpretation of these changes in terms of the measurability of such quantities? (they were measurable before the change, no more

property values are usually hypothesized to be real numbers, or however, when the previous consideration is taken into account, rationals, and therefore always scalars. On the other hand, if it is recognized that (real) numbers are linguistic means to express our knowledge, then the conclusion should be drawn that scalars are only one possible choice to formalize property values, so that other options, such as intervals, probability distributions, fuzzy subsets, could be adopted. Apart from tradition, I suggest that a single, but fundamental, reason remains today explaining why property values are so commonly expressed by scalars: such values act as input data in inferential structures, as it is the case of physical laws, which are deemed to deal with scalars. Indeed, no logical reasons prevent to express inferences, such as the above mentioned "the acceleration generated on a body with mass m by a force F is equal to F/m", in terms of non-scalar values, e.g., intervals or fuzzy subsets, provided that the functions appearing in such inferences, the ratio in this case, are properly defined for these non-scalar values.

5.5. Balancing specificity and trust: a pragmatic choice

The analysis stage does not univocally determine the form of measurement results also because *quality of measurement is a multi-dimensional characteristic*. According to Bertrand Russell: "all knowledge is more or less uncertain and more or less vague. These are, in a sense, opposing characters: vague knowledge has more likelihood of truth than precise knowledge, but is less useful. One of the aims of science is to increase precision without diminishing certainty" [Russell 1926]. *The same empirical knowledge available on a measurand value can be formally expressed by balancing two components*:

- one defining the specificity of the value: sometimes this component is called *precision* or, at the opposite, *vagueness*;
- one stating the trust attributed to it: neither *accuracy* nor *trueness* (the latter term is used in [ISO 1998], but not in the VIM [ISO 1993]) have been mentioned here. If a reference value is not known, such quantities are simply undefined; in the opposite case, accuracy can be thought of as the subject-independent version of trust.

Until the available knowledge on the measurand is not experimentally enhanced, if one component is increased the other one should be decreased. An instance of such a trade-off is the probabilistic relation between confidence intervals and confidence levels. In the general case, the measurement result could be expressed as, e.g., a fuzzy subset with an associated possibility measure / distribution (see at this regards for example [Benoit et al 2005]), so that the assignment of the two components stating the quality of measurement remains largely a task based on the experience of

after; they have never been actually measurable; ...). In more general terms, from the fact that any physical measuring system has a finite resolution the conclusion follows that all measurement results must always be expressed as discrete (and actually with a small number of significant digits) entities: does it imply according to the realist view that 'real' measurements are only approximations of 'ideal' measurements, or what else?" [Mari 2005].

the subject.
From this point of view *the approach followed by the GUM* [ISO 1995] *is hybrid*, being based on two complementary models, both of them probabilistic in their bases but opposite in the component of quality they emphasize (I should point out that the following analysis presents my viewpoint on the GUM, and is not literally faithful to the GUM itself, which is however repeatedly quoted henceforth; a good and synthetic (and faithful) synthesis of the GUM is [Taylor, Kuyatt 1994]). Let us call them "primary" model and "secondary" model respectively.
The *primary model* assumes that the measurement result is expressed as a couple:

<estimated measurand value, standard uncertainty>

where the first term is a scalar and the second one is interpreted as a standard deviation of the estimated measurand value, either derived from an experimental frequency distribution or obtained from an assumed underlying probability density function (7). In the first case the estimated measurand value is defined as the mean value of the experimental set $\{v_i\}$, i=1,...,n:

$$\bar{v}=\frac{1}{n}\sum_{1}^{n} v_i$$

and the *standard uncertainty* is defined as the experimental standard deviation of the mean value:

$$u(\bar{v})=s(\bar{v})=\sqrt{\frac{1}{n(n-1)}\sum_{1}^{n}(v_i-\bar{v})^2}$$

If an experimental frequency distribution is not available, the standard uncertainty "is evaluated by scientific judgment based on all the available information on the possible variability" of the property (as examples of such information sources, the GUM mentions the following: "previous measurement data; experience with or general knowledge of the behavior and properties of relevant materials and instruments; manufacturer's specifications; data provided in calibration and other certificates; uncertainties assigned to reference data taken from handbooks." The focus on physical quantities is here manifest). The choice of assuming a maximally specific value for the measurand implies that its quality is entirely expressed as trust, by means of the standard uncertainty.
The *secondary model* assumes the measurement result to be expressed as an interval, whose half

7 This double option highlights, once more, the pragmatic orientation of the GUM: while traditional distinctions are aimed at identifying "types" of uncertainty (or of error, of course, as in the case of random vs. systematic error), thus assuming an ontological basis for the distinction itself, the GUM distinguishes between methods to evaluate uncertainty. Furthermore, the GUM removes any terminological interference by adopting a Recommendation issued by the International Committee for Weights and Measures (CIPM) in 1980 and designating as "Type A" the evaluations performed "by the statistical analysis of series of observations", and as "Type B" the evaluations performed "by other means". The GUM itself stress then that "the purpose of the Type A and Type B classification is to indicate the two different ways of evaluating uncertainty components and is for convenience of discussion only; the classification is not meant to indicate that there is any difference in the nature of the components resulting from the two types of evaluation".

width is called *expanded uncertainty*, U, and is derived from the primary model by multiplying the standard uncertainty $u(\bar{v})$ by a positive coefficient k, called "coverage factor", typically in the range 2 to 3, $U=k\,u(\bar{v})$. Such an interval, $\underline{v}=[\bar{v}-U, \bar{v}+U]$, has the goal "to encompass a large fraction of the distribution of values that could reasonably be attributed to the measurand". The quality of the value is now in principle formalized in terms of both specificity and trust, as related respectively to the expanded uncertainty and the encompassed "fraction of the distribution", interpreted as a probability measure and called "level of confidence" of the interval. On the other hand, since "it should be recognized that in most cases the level of confidence (especially for values near 1) is rather uncertain" and therefore difficult to assign, the standardized decision is made of choosing a level of confidence above 0,95, by suitably increasing the expanded uncertainty as believed to be required: the expanded uncertainty, and therefore the specificity, is thus in practice the only component which expresses the quality of measurement.

The reasons of this double modeling are explicitly pragmatic:

- the primary model is aimed at propagating uncertainties through functional relationships;
- the secondary model is aimed at comparing property values to ascertain whether they are compatible to each other.

Let us introduce the main features of these application categories.

5.6. Propagation of uncertainty

Let q be a property computed by a function f, $q=f(p_1,p_2,\ldots,p_k)$, where each p_j is a property whose value is assumed to be available (because either measured or in its turn computed) but uncertain, and the analytical form of the function f is assumed to be known. If the information on each input property p_j is expressed as a couple $<\bar{v}_j, u(\bar{v}_j)>$, i.e., according to the primary model, and the information on the output property q must also be expressed as a couple $<\bar{v}_q, u(\bar{v}_q)>$, then the problem is to derive $<\bar{v}_q, u(\bar{v}_q)>$ from the set $\{<\bar{v}_j, u(\bar{v}_j)>\}_j$. Since the input information is uncertain and the sought output information is also expected to be uncertain, this derivation problem is called *propagation of uncertainty*. Such a problem has been traditionally applied in derived measurement, in which the input properties are experimentally measured and the function f formalizes a (physical) law. On the other hand, the uncertainty should also be propagated when the dependence of a measurand from one or more influence quantities is analytically known and both the measurand and its influence quantities are expressed as uncertain, the output quantity being in this case the measurand specified in reference to the identified influence quantities. This shows the generality of the problem. The choice of expressing the measurement result for each property p_j as a couple $<\bar{v}_j, u(\bar{v}_j)>$ allows to compute the output property value $\bar{v}_q$ and its standard uncertainty $u(\bar{v}_q)$ by means of separate procedures:

- $\bar{v}_q$ is obtained by applying the function *f* to the estimated values of the input properties: $\bar{v}_q = f(\bar{v}_1, \bar{v}_2, \ldots, \bar{v}_k)$;
- $u(\bar{v}_q)$ is obtained by assuming that the uncertainty on each property p_j produces a deviation Δv_j from the mean value $\bar{v}_j$, so that the problem is to derive the standard deviation of $\bar{v}_q$ from $f(\bar{v}_1 + \Delta v_1, \bar{v}_2 + \Delta v_2, \ldots, \bar{v}_k + \Delta v_k)$.

The technique recommended by the GUM is based on the hypothesis that the function *f* can be approximated by its Taylor series expansion in the *k*-dimensional point $< \bar{v}_1, \bar{v}_2, \ldots, \bar{v}_k >$. In the simplest case, in which all input properties are independent and *f* is "linear enough" around this point, the series expansion can be computed up to the first-order term:

$$u^2(\bar{v}_q) = \sum_{j=1}^{k} c_j^2 u^2(\bar{v}_j)$$

an expression called *law of propagation of uncertainty*, which shows that the standard uncertainty for the output property, called "combined standard uncertainty" by the GUM, depends on the weighted quadratic sum of the standard uncertainties of the input properties. Each weight c_j:

$$c_j = \left.\frac{\partial f}{\partial p_j}\right|_{\bar{v}_1, \bar{v}_2, \ldots, \bar{v}_k}$$

i.e., the partial derivative of the function *f* with respect to the *j*-th property as computed in the point $< \bar{v}_{1,} \bar{v}_{2,} \ldots, \bar{v}_k >$, operates as a "sensitivity coefficient". This dependence becomes more and more complex as higher-order terms in the Taylor series expansion and / or the correlations among the input properties are taken into account: at this regards some further technical considerations can be found in the GUM, and several cogent examples are presented in [Lira 2002].

This logic of solution to the problem of the propagation of uncertainty is based on the traditional choice of expressing the property value as a scalar entity, distinct from the parameter specifying its quality: property values are dealt with in a deterministic way and analytical techniques are applied for formally handling the uncertainty (8).

5.7. Comparison of uncertain property values

I have proposed above an operational definition of properties based on the empirical substitutability of objects: if two objects x_1 and x_2 are recognized as mutually substitutable for some purpose, then

8 The working group who created the GUM is currently preparing some addenda to it, and in particular the "Supplement 1: Numerical methods for the propagation of distributions", which presents an alternative solution to the problem of uncertainty propagation. Whenever input property values can be expressed as probability density functions, the whole functions can be propagated, to obtain a "combined propagated function". This logic is in principle more general than the one endorsed by the GUM, since the mean value and its standard deviation are trivially derived from a probability density function. On the other hand, since the combined propagated function cannot generally be obtained by analytical techniques, the propagation can be performed in a numerical way, typically by the Monte Carlo method.

there must exist a property p such that $p(x_1)=p(x_2)$. On the other hand, whenever the property values are recognized to be uncertain (and, for example, are expressed as $<\bar{v}, u(\bar{v})>$) such an equality at the same time:

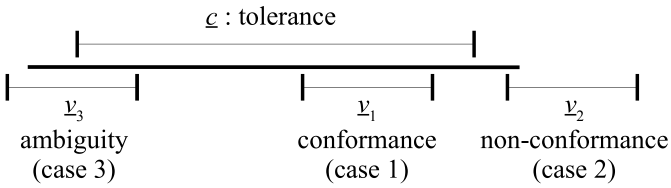
- is ambiguous, since it is not clear whether it requires that $\bar{v}_1=\bar{v}_2$ independently of their uncertainties, or also that $u(\bar{v}_1)=u(\bar{v}_2)$;
- constitutes a too narrow constraint, since mutual substitutability is guaranteed also in the case $\bar{v}_1$ and $\bar{v}_2$ are "close enough" relatively to their uncertainties, even if not identical.

When uncertainty is taken into account, mutual substitutability *does not require the equality of property values, but more generally their "compatibility"*. More than by means of sophisticated analytical techniques, the check of compatibility between property values is customarily performed by their direct comparison. To this goal, the simplest solution is to express property values as intervals, as in the secondary model, and to formalize their compatibility in terms of their set-theoretical intersection, which must be non-null. The GUM itself acknowledges the practical scope of the secondary model, introduced "to meet the needs of some industrial and commercial applications, as well as requirements in the area of health and safety". Indeed, this check of compatibility has at least two general applications:

- *it is a means to obtain information on the repeatability of a measurement*: while consecutive measurements produce compatible results, the inference can be drawn that the object under measurement is not changed with respect to the measurand;
- *it is a means to decide about the conformance with given specifications in presence of uncertainty*: this pragmatic issue is so important that deserves some further consideration.

A technical specification on a property is usually expressed as an interval of conformance $\underline{c}$, also called "tolerance", i.e., the subset of the property values which are considered acceptable for the given application. If also measurement results for that property are expressed as intervals $\underline{v}$, the decision on the conformance of $\underline{v}$ with the specification formalized by $\underline{c}$ requires the comparison of the two intervals. The outcome can be (see [ISO 1998b]):

- (case 1) if the property value is completely within the tolerance, $\underline{v}\subset\underline{c}$, then the value is assumed to be in conformance with the specification;
- (case 2) else if the property value is completely outside the tolerance, $\underline{v}\cap\underline{c}=\emptyset$, then the value is assumed not to be in conformance with the specification;
- (case 3) otherwise, i.e., if the property value is only partially within the tolerance, $\underline{v}\cap\underline{c}\neq\emptyset$ but not $\underline{v}\subset\underline{c}$, then the situation is ambiguous.



Whenever the property values are expressed with a non-null uncertainty, case 3 of ambiguity can always appear in the borderline situations. On the other hand, the frequency of this case statistically decreases as the width of the interval $\underline{v}$ decreases (whereas in the extreme situation in which the width of $\underline{v}$ is greater than the width of $\underline{c}$ the first case cannot occur): *the quality of measurement influences the ambiguity of the conformance decision*.

5.8. Uncertainty evaluation as a pragmatic decision

In presence of ambiguity, a decision can be made only on the basis of a contextual criterion. If, for example, the conformance to a specification is the condition required by a subject A (the buyer, the evaluator) to accept an entity (a product, a service, ...) produced by a subject B (the seller, the maker), then two positions can be assumed:

- "in defense of buyer", ambiguity is dealt with as non-conformance, i.e., *in doubt refuse*;
- "in defense of seller", ambiguity is dealt with as conformance, i.e., *in doubt accept*.

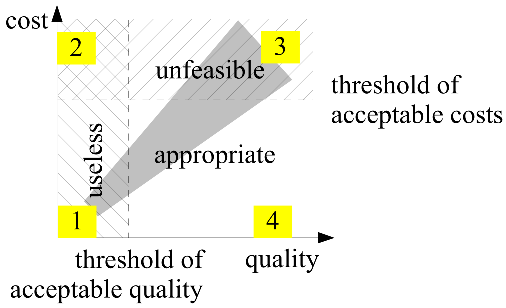
This alternative maps the traditional distinction between the so called "type 1" errors (wrong acceptation) and "type 2" errors (wrong refusal). Under the hypotheses that (1) the object state can be expressed in dichotomic way, in terms of either conformance or non-conformance, and that (2) the decision can only be either "accept" or "refuse", four situations can occur:

		Object state	
		non-conformance	conformance
Decision	refuse	ok: correct refusal	type 2 error
	accept	type 1 error	ok: correct acceptation

A further position is in principle possible: if measurement results can be obtained with a reduced uncertainty, the ambiguity could be removed by reclassifying case 3 as either case 1 or case 2. This option highlights the pragmatic nature of the evaluation of measurement quality: since enhancing the measurement quality generally implies increasing the costs of the resources required to perform the measurement process, the issue arises of *balancing the costs of such resources with the quality of the measurement results*. The stated goal of the process should allow to identify a lower bound for acceptable quality and an upper bound for acceptable costs, so that the decision space can be split in three subspaces, for decisions leading respectively to:

- *useless* measurements, which, independently of their costs, have a quality insufficient with respect to the given goal (in conformity decision this corresponds to case 3);
- *unfeasible* measurements, which, independently of their quality, have costs unacceptable with respect to the given goal;
- *appropriate* measurements, for which the trade-off quality vs. costs is compatible with the given goal.

Accordingly, the decision should be made before measurement about its minimum acceptable quality threshold, expressed by the so called *target uncertainty*, so that the measurement process should be performed according to the following procedure:



1. decide the minimum acceptable quality, i.e., the target uncertainty, u_T, and the maximum acceptable costs, i.e., the resource budget (i.e., define the four subspaces in the previous chart);
2. estimate the minimum costs required to obtain u_T: if such costs are beyond the resource budget, then stop as unfeasible measurement (the procedure stops in subspaces 2 or 3);
3. identify the components which are deemed to be the main contributions to the uncertainty budget;
4. choose an approximate method to combine such contributions, credibly leading to overestimate the combined uncertainty;
5. perform the measurement by keeping into account the identified contributions and evaluate the result by combining them, thus obtaining a measurement uncertainty u_M;
6. compare u_M to u_T : if $u_M < u_T$, then exit the procedure by stating the obtained data as the result of an appropriate measurement (area 4);
7. estimate the current costs: if such costs are beyond the resource budget, then stop as unfeasible measurement (area 2);
8. identify further contributions and / or enhance the method to combine them;
9. repeat from 5.

The pragmatic ground of this algorithm is manifest: as soon as the available information allows to unambiguously satisfy the goal for which the measurement is performed, the process should be stopped. Any further activity is not justified, because useless and uselessly costly: *the concept of "true uncertainty" is simply meaningless*.

6. Conclusions

The traditional concept of measurability is grounded in ontology: each specific property, such as “the length of this table”, is measurable because it inherently has a “true value”, whose determination is the aim of measurement, so that the empirical inability of obtaining such true values is accounted for as caused by errors in the measurement process. The concept of measurability presented in this paper is instead a pragmatic one: measurement results must be assigned (and not determined) according to the goals for which the measurement is performed, with the consequence that they are adequate if they meet such goals. Measurement results are evaluated and formally expressed by suitably eliciting the information on the measurand value and its quality experimentally acquired in the measurement process. In this evaluation a critical role is played by the subject; indeed, as the GUM considers, no method for evaluating the quality of measurement can be a “substitute for critical thinking, intellectual honesty, and professional skill; (...) the quality and utility of the uncertainty quoted for the result of a measurement ultimately depends on the understanding, critical analysis, and integrity of those who contribute to the assignment of its value”. In this perspective, intersubjectivity and objectivity become the pragmatic target of measurement, instead of its preliminary ontological conditions. Measurement is recognized to be a model-based process, and thus it is emphasized its dependence to an interpretive activity that is pre-empirical: scientifically organized bodies of knowledge, modeling methodologies, formally-defined constraints, physical instrumentations, ethical responsibility, all contribute to the social value of measurement even in the current epistemologically relativistic world.

References

[Benoit et al 2005] E.Benoit, L.Foulloy, G.Mauris, “Fuzzy approaches for measurement”, in: P.Sydenham, R.Thorn (eds.), Handbook of Measuring System Design, pp.60-67, Wiley, 2005 – ISBN 0-470-02143-8.

[Boumans 2007] M.Boumans, “Invariance and calibration”, *in this book*.

[Bridgman 1927] P.W.Bridgman, “The Logic of Modern Physics”, MacMillan, New York, 1927.

[Carbone et al. 2006] P.Carbone, L.Buglione, L.Mari, D.Petri, “Metrology and software measurement: a comparison of some basic characteristics”, Proc. IEEE IMTC 2006, pp.1082-1086, Sorrento, 24-27 Apr 2006.

[D’Agostini 2003] G.D’Agostini, “Bayesian reasoning in data analysis - A critical introduction”, World Scientific Publishing, Singapore, 2003.

[Finkelstein, Leaning 1984] L.Finkelstein, M.Leaning, “A review of the fundamental concepts of measurement”, Measurement, 2, 1, 1984.

[Hacking 1983] I.Hacking, “Representing and intervening - Introductory topics in the philosophy of natural science”, Cambridge University Press, Cambridge, 1983.

[ISO 1993] International Organization for Standardization, “International vocabulary of basic and general terms in metrology”, second edition, Geneva, 1993 (published by ISO in the name of BIPM, IEC, IFCC, IUPAC, IUPAP and OIML).

[ISO 1995] International Organization for Standardization, “Guide to the expression of uncertainty in measurement”, Geneva, 1993, amended 1995 (published by ISO in the name of BIPM, IEC, IFCC, IUPAC, IUPAP and OIML)

(also ISO ENV 13005: 1999).
[ISO 1998] International Organization for Standardization, "Accuracy (trueness and precision) of measurement methods and results – Part 1: General principles and definitions", Geneva, 1998.
[ISO 1998b] International Organization for Standardization, 14253-1: "Geometrical Product Specification - Inspection by measurement of workpieces and measuring instruments". Part 1: "Decision rules for proving conformance or non-conformance with specification", Geneva, 1998.
[Kuhn 1970] T.S.Kuhn, "The structure of scientific revolutions", University of Chicago Press, Chicago, 1970.
[Lira 2002] I.Lira, "Evaluating the measurement uncertainty – Fundamentals and practical guidance", Institute of Physics, 2002 – ISBN 0-750-30840-0.
[Mari 1997] L.Mari, "The role of determination and assignment in measurement", Measurement, 21, 3, pp.79-90, 1997.
[Mari 1999] L.Mari, "Notes towards a qualitative analysis of information in measurement results", Measurement, 25, 3, pp.183-192, 1999.
[Mari 2000] L.Mari, "Beyond the representational viewpoint: a new formalization of measurement", Measurement, 27, 1, pp.71-84, 2000.
[Mari 2003] L.Mari, "Epistemology of measurement", Measurement, 34, 1, pp.17-30, 2003.
[Mari 2005] L.Mari, "The problem of foundations of measurement", Measurement, 38, 4, pp.259-266, 2005.
[Mari 2005b] L.Mari, "Explanation of key error and uncertainty concepts and terms", in: P.Sydenham, R.Thorn (eds.), Handbook of Measuring System Design, pp.331-335, Wiley, 2005 – ISBN 0-470-02143-8.
[Mari, Zingales 2000] L.Mari, G.Zingales, "Uncertainty in Measurement Science", in: Measurement Science – A discussion, K.Karija, L.Finkelstein (eds.), Ohmsha, IOS Press, 2000 - ISBN 4-274-90398-2 (Ohmsha Ltd.) / 1-58603-088-4 (IOS Press).
[Michell 2007] J.Michell, "Representational theory of measurement", *in this book*.
[Morgan 2007] M.S.Morgan, "A brief and illustrated analytical history of measuring in Economics", *in this book*.
[Phillips et al. 2001] S.D.Phillips, W.T.Estler, T.Doiron, K.R.Eberhardt, M.S.Levenson, "A careful consideration of the calibration concept", Journal of Research of the National Institute of Standards and Technology, 106, 2, pp.371-379, 2001 (available online: http://www.nist.gov/jresp).
[Quine 1966] W.V.O.Quine, "The scope and language of science", in: The ways of paradox, Random House, 1966.
[Rossi 2006] G.B.Rossi, "An attempt to interpret some problems in measurement science on the basis of Kuhn's theory of paradigms", Measurement, 39, 6, pp.512-521, 2006.
[Russell 1926] B.Russell, "Theory of Knowledge", in: The Encyclopedia Britannica, 1926.
[Shannon 1948] C.E.Shannon, "A mathematical theory of communication", Bell System Technical Journal, 27, p. 379-423 and 623-656, 1948.
[Taylor, Kuyatt 1994] B.N.Taylor, C.E.Kuyatt, "Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results", National Institute of Standards and Technology Technical Note 1297, 1994 (available online: http://physics.nist.gov/Pubs/guidelines/TN1297/tn1297s.pdf).